Université de Montréal

Guider les thérapeutes dans l'amélioration de la réponse thérapeutique d'un patient à l'aide de l'intelligence artificielle pour la thérapie par Avatar

Par

Alexandre Hudon

Département de psychiatrie et d’addictologie, Faculté de Médecine

Thèse présentée en vue de l’obtention du grade de doctorat

en psychiatrie et addictologie

Janvier 2024

Université de Montréal

Département de psychiatrie et d’addictologie, Faculté de Médecine

---

Cette thèse intitulée

Guider les thérapeutes dans l'amélioration de la réponse thérapeutique d'un patient à l'aide de l'intelligence artificielle pour la thérapie par Avatar

Présenté par
Alexandre Hudon

A été évalué(e) par un jury composé des personnes suivantes

Frédérick Aardema
Président-rapporteur

Alexandre Dumais
Directeur de recherche

Stéphane Potvin
Codirecteur

Jean-Philippe Miron
Membre du jury

Ridha Joober
Examinateur externe

Résumé

Dans le contexte de la psychiatrie moderne, les troubles de santé mentale graves et persistants, notamment la schizophrénie, présentent des défis thérapeutiques considérables. La schizophrénie, en particulier, impacte profondément le fonctionnement des individus, notamment à travers des symptômes tels que les hallucinations. Ces manifestations peuvent affecter de manière délétère les interactions interpersonnelles des patients, leur régulation émotionnelle et sont associées à un taux de suicide nettement supérieur à celui de la population générale. Face à la résistance aux traitements psychopharmacologiques de première ligne, des thérapies non-pharmacologiques ont été développées, dont la thérapie par Avatar (TA) en réalité virtuelle (RV), qui permet aux patients souffrant d'hallucinations auditives résistantes aux médicaments d'interagir avec une représentation en 3D de leur voix perturbatrice. Bien que cette thérapie ait montré une réduction des hallucinations et une amélioration de la qualité de vie, certains patients restent réfractaires, ce qui pose des défis dans leur prise en charge.

L'utilisation de données quantitatives en médecine a amélioré les approches thérapeutiques pour les patients présentant des troubles complexes. Les thérapeutes s'appuient sur des outils numériques basés sur des algorithmes mathématiques pour guider leur prise de décision clinique. L'intelligence artificielle, notamment l'apprentissage machine, se développe en médecine clinique pour aider les thérapeutes à prédire l'évolution des patients et à choisir le traitement approprié. Cependant, en psychiatrie, où les modèles biopsychosociaux sont prédominants, l'utilisation de données quantitatives dans les interventions psychothérapeutiques est moins courante. Les thérapeutes s'appuient souvent sur des guides de pratique génériques et leur expérience clinique, ce qui peut être insuffisant, particulièrement pour les patients atteints de schizophrénie réfractaire à la médication.

Cette thèse vise à intégrer les principes de l'intelligence artificielle dans la thérapie par Avatar pour améliorer la réponse thérapeutique des patients souffrant de schizophrénie avec des hallucinations auditives réfractaires aux médicaments. Elle se concentre sur cinq sous-objectifs : l'utilisation de l'intelligence artificielle pour prédire les issues cliniques des patients atteints de

troubles mentaux graves, l'évaluation des possibilités d'intégration de l'IA en psychothérapie, l'utilisation de l'apprentissage machine dans la TA pour l'annotation automatique des séances, l'intégration des algorithmes d'apprentissage machine pour prédire la réponse des patients à la TA, et l'identification des facteurs prédictifs multimodaux améliorant la prédiction des issues cliniques.

La première partie de la thèse concerne une étude transversale utilisant l'apprentissage machine pour identifier des facteurs prédicteurs de violence chez des patients souffrant de troubles mentaux graves, en se concentrant sur les liens entre la consommation de cannabis et la violence. Les résultats indiquent que l'utilisation du cannabis est un prédicteur clé de la violence.

La deuxième partie évalue l'intégration de l'IA en psychothérapie, à travers deux revues de littérature. La première examine l'utilisation de l'apprentissage machine sur des petites bases de données pour l'annotation automatisée des transcriptions de séances thérapeutiques. La seconde se concentre sur l'utilisation de l'IA, notamment les réseaux neuronaux, en psychothérapie clinique.

La troisième partie explore l'utilisation de l'apprentissage machine dans la TA, en comparant différents algorithmes pour la classification automatisée des interactions thérapeutiques et en utilisant des techniques d'apprentissage non-supervisé pour identifier des groupes d'interactions.

La quatrième partie s'appuie sur les précédentes pour intégrer des algorithmes d'apprentissage machine afin de prédire la réponse clinique des patients à la TA, en se basant sur les interactions thérapeutiques lors de la première séance de thérapie immersive.

Enfin, la cinquième partie présente deux études supplémentaires : une sur l'identification des émotions dans la TA et une autre sur le concept de dyades thérapeutiques, illustrant l'importance de la relation entre le thérapeute et le patient dans la TA.

En conclusion, cette thèse apporte une contribution significative à la compréhension des processus thérapeutiques dans la TA via l'intelligence artificielle, en ouvrant la voie à des approches personnalisées basées sur des modèles mathématiques pour améliorer la prise en

charge des patients atteints de schizophrénie réfractaire.

Mots-clés : Apprentissage machine, Thérapie Avatar, Thérapie assistée par la Réalité virtuelle, Schizophrénie, Analyse de contenu, Psychothérapie, Réalité virtuelle

Abstract

In the context of modern psychiatry, severe and persistent mental health disorders, particularly schizophrenia, present significant therapeutic challenges. Schizophrenia profoundly impacts the functioning of individuals, especially through symptoms like hallucinations. These manifestations can detrimentally affect the interpersonal interactions of patients, their emotional regulation, and are associated with a suicide rate significantly higher than that of the general population. Faced with resistance to first-line psychopharmacological treatments, non-pharmacological therapies have been developed, including Avatar Therapy (TA) in virtual reality (RV), which allows patients suffering from medication-resistant auditory hallucinations to interact with a tri-dimensional representation of their most disturbing voice. Although this therapy has shown a reduction in hallucinations and an improvement in the quality of life, some patients remain refractory, posing challenges in their management.

The use of quantitative data in medicine has improved therapeutic approaches for patients with complex disorders. Therapists rely on digital tools based on mathematical algorithms to guide their clinical decision-making. Artificial intelligence, particularly machine learning, is currently developing in clinical medicine to help therapists predict patients' outcomes and choose appropriate treatments. However, in psychiatry, where biopsychosocial models are predominant, the use of quantitative data in psychotherapeutic interventions is less common. Therapists often rely on generic practice guidelines and their clinical experience, which may be insufficient, especially for patients with medication-refractory schizophrenia.

This thesis aims to integrate the principles of artificial intelligence into Avatar Therapy to improve the therapeutic response of patients suffering from schizophrenia with medication-resistant auditory hallucinations. It focuses on five sub-objectives: using artificial intelligence to predict clinical outcomes of patients with severe and persistent mental disorders, evaluating the possibilities of integrating AI in psychotherapy, using machine learning in AT for the automated annotation of sessions, integrating machine learning algorithms to predict patients' responses to

AT, and identifying multimodal predictive factors to improve the prediction of clinical outcomes.

The first part of the thesis involves a cross-sectional study using machine learning to identify predictors of violence in patients with severe and persistent mental disorders, focusing on the links between cannabis use and violence. The results indicate that cannabis use is a key predictor of violence. The second part assesses the integration of AI in psychotherapy through two literature reviews. The first review examines the use of machine learning on small datasets for the automated annotation of therapeutic session transcripts. The second focuses on the use of AI, particularly neural networks, in clinical psychotherapy.

The third part explores the use of machine learning in AT, comparing different algorithms for the automated classification of therapeutic interactions and using unsupervised learning techniques to identify groups of interactions. The fourth part builds on the previous ones to integrate machine learning algorithms to predict the clinical response of patients to AT, based on the therapeutic interactions during the first immersive therapy session. Finally, the fifth part presents two additional studies: one on identifying emotions in AT and another on the concept of therapeutic dyads, illustrating the importance of the relationship between the therapist and the patient in AT.

In conclusion, this thesis makes a significant contribution to understanding the therapeutic processes in AT through artificial intelligence, paving the way for personalized approaches based on mathematical models to improve the management of patients with medication-refractory schizophrenia.

Keywords : Machine learning, Avatar Therapy, Virtual Reality-Assisted Therapy, Schizophrenia, Content Analysis, Psychotherapy, Virtual Reality.

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles et abréviations

ADN: Acide désoxyribonucléique

AT : Avatar Therapy

AVH: Auditory verbal hallucinations

CBT : Cognitive Behavioral Therapy

CBTp : Cognitive Behavioral Therapy for psychosis

DSM -5: Manuel diagnostique et statistique des troubles mentaux

ECR : Essai contrôlé randomisé

GWAS : Genome-Wide Association Studies

HAV : Hallucinations auditives verbales

ICM-10 : Classification internationale des maladies 10ième édition

PSYRATS-HA : Échelle des hallucinations auditives du *Psychotic Symptoms Rating Scale*

RCT : Randomized controlled trial

RV : Réalité virtuelle

SMD : Severe mental disorder

SNP : polymorphisme nucléotidique simple

SRT : Schizophrénie résistante au traitement

TA : Thérapie Avatar

TCC: Thérapie Cognitivo-Comportementale

TCCp : Thérapie Cognitivo-Comportementale pour la psychose

TRS : Treatment resistant schizophrenia

VRT : Virtual Reality assisted Therapy

VR : Virtual reality

Cette thèse est dédiée à toutes les patientes et tous les patients pour lesquels la médecine moderne n'a pas encore su aider. Nous allons finir par y arriver ensemble.

Remerciements

Je ressens une profonde gratitude envers plusieurs personnes clés qui, chacune à leur manière, m'ont apporté leur aide pour me permettre de croître sur le plan personnel et professionnel :

À mon directeur et mon co-directeur de recherche qui m’ont soutenu tout au long de cette aventure. Vous avez cru en mes capacités et vous m’avez donné tout le nécessaire afin que je puisse m’épanouir dans le monde de la recherche et intégrer mes compétences en ingénierie logicielle dans le domaine de la psychiatrie clinique.

- Dr Alexandre Dumais : Je tiens à vous remercier pour l'excellent modèle de rôle que vous avez été pour moi durant tout ce parcours. Vous m'avez donné le goût de me dépasser au quotidien en jumelant la recherche et la clinique tout en maintenant un équilibre de vie. Ce fût un grand plaisir pour moi d'avoir pu apprendre la TA à vos côtés, d'intégrer ce monde intéressant qu'est la fusion entre la réalité virtuelle et la psychiatrie. Je vous remercie pour votre confiance et tout le support que vous m'avez prodigué pour que je puisse passer à travers les obstacles rencontrés et puisse m'épanouir à travers diverses opportunités pédagogiques, cliniques et d'érudition.

- Dr. Stéphane Potvin : Ce fût un grand plaisir pour moi de vous avoir comme co-superviseur. Votre expérience sur le terrain, vos conseils, votre temps et surtout le contact très agréable m’a permis d’aller plus loin dans mes réflexions et dans la réalisation de mes travaux de recherche. Les commentaires constructifs soutenus tout au long de mes travaux m’ont permis de toujours fournir un contenu de qualité plus élevé lors de la présentation de mes travaux. Votre zenitude m’a aussi inspiré à garder mon calme lorsque les choses ne vont pas dans la direction espérée et de se mettre en mode solutions plutôt qu’en mode rumination.

À toute l’équipe du laboratoire Avatar, j’ai adoré travailler au sein de cette communauté qui adopte un mode familial plutôt que compétitif, ce qui rend la recherche agréable et ouvre la possibilité de coconstruire des projets intéressants en s’appuyant sur nos forces mutuelles.

- Sabrina Giguère : Ta bonne humeur est contagieuse ainsi que ton approche auprès des patients. J'ai adoré travailler avec toi et te côtoyer avant et après les séances de thérapie, de discuter des embûches du parcours et comment les surmonter.
- Mélissa Beaudoin : Ce fût un réel plaisir d'évoluer dans ce parcours à tes côtés. Merci d'avoir toujours été partante pour des projets de fou avec un court échéancier. Tes rétroactions constructives ont toujours été très apprécié et m'ont permis d'aller plus loin dans mes réflexions.
- Dr. Laura Dellazizzo : Ton expertise en recherche, tes rétroactions constructives, ton aide sur les différents projets (et j'en passe) m'ont permis de passer à travers les différents défis que m'ont offert ce parcours.
- Stéphane Gagnon : Merci beaucoup d'avoir pris le temps de me partager ton expertise dans le cadre de la TA. Tes techniques relationnelles sont inspirantes et ce fût très inspirant pour moi de voir la relation thérapeutique que tu mets sur pied avec tes patients.
- Marie-Andrée Lapierre : Ce fût toujours très agréable de te rencontrer. Merci pour le soutien et le support pour le recrutement et le suivi lors des thérapies.
- Kingsada Phraxayavong : Je pense que je ne sais même pas par où commencer ce remerciement. Tu as été pour moi une source bienveillante de motivation, d'inspiration et tu m'as permis d'aller de l'avant avec l'ensemble de mes projets tout en fournissant tout le nécessaire pour que ceux-ci réussissent. Ton organisation légendaire ainsi que ta disponibilité m'a permis de traverser cette étape avec tout ce dont j'avais besoin pour évoluer favorablement. C'est un réel plaisir de collaborer avec toi.
- Jonathan Couture, Nayla Léveillé, Sabrina Quilliam, Veronica Iammateo, Sophie Rodrigues-Coutlée, Alexandre Fortier, Dre. Katerina Sanchez-Schicharew, Simon-Pierre Bernard Arevalo: Ce fût très agréable de collaborer avec vous sur les différents projets de la TA. Merci pour votre aide dans la réalisation de cette étape de mon parcours.

À la direction du programme de résidence en psychiatrie et la direction du département universitaire qui m'ont permis de jumeler une résidence en psychiatrie avec un doctorat de troisième cycle en toute simplicité.

- Dr. François Lespérance : Vous avez été présent tout au long de mon parcours pour me soutenir et m'encourager dans mes démarches en m'assurant d'avoir les contacts et les ressources nécessaires pour évoluer favorablement à travers ce parcours. Vous avez cru en mes capacités et m'avez permis de me dépasser durant cette aventure.

- Dr. Yvan Pelletier : Vos multiples appuis et les contacts toujours des plus agréables et soutenant m'ont permis de bénéficier à la fois des apprentissages cliniques, en recherche et en leadership. Ceci m'a permis, je crois, de tirer le meilleur de mon programme de résidence en sachant que vous avez cru en mes capacités et en me supportant tout au long de celui-ci.

- Dr. Lionel Cailhol : Rapidement dans mon parcours, vous m'avez donné les outils et les opportunités de développer mon intérêt pour la recherche et ses applications cliniques. Votre soutien et votre ouverture à ce que je puisse m'épanouir au sein de votre établissement me permettront d'aller encore plus loin dans mon plan de carrière afin de me réaliser à titre de clinicien-chercheur.

- Dre. Stéphanie Borduas Pagé : Tu m'as intégré dans toutes les sphères de ta pratique clinique et m'as permis de mettre sur pied des projets afin que je puisse développer mes habiletés de supervision au sein du merveilleux service de psychiatrie légale de l'Institut universitaire de santé mentale de Montréal. Ton approche avec les patients et les apprenants est inspirant. Ce fût un grand plaisir pour moi de t'avoir comme modèle de rôle inspirant et j'ai bien hâte de te rejoindre à titre de collègue dans un avenir très proche.

À ma conjointe Dre. Caroline Gaudreau-Ménard, qui a toujours été à mes côtés pour me soutenir dans mes projets de fou et qui a été de bons conseils lors de périodes de stress. Tu es une femme extraordinaire aux multiples qualités qui a à cœur ses patients, ses amies et ses proches et ce côté humain et d'une grande inspiration.

À ma famille Julien, Diane, Marc-Antoine et Nicolas, et mon parrain Claude, et ma belle-famille (Marie, Marc, Élisabeth, Raymond et la petite Rose), vous êtes une source de motivation pour moi d'aller plus loin dans mon développement personnel. Merci de me soutenir dans tous mes projets et mes idées et d'avoir mis tout ce dont j'avais besoin à ma disposition pour réussir à travers mon parcours académique et personnel.

Finalement, je tiens à remercier les différentes instances financières qui m'ont épaulé durant ce parcours, me permettant d'avoir les ressources nécessaires pour développer mes projets : IVADO et l'Université de Montréal.

Chapitre 1 – Introduction

En 2024, dans le domaine de la psychiatrie, plusieurs psychopathologies complexes demeurent à l'étude et méritent d'être davantage explorées sur le plan clinique et thérapeutique. À cet effet, la schizophrénie, du fait de son spectre clinique et de ses variantes de réponses au traitement, représente un défi de taille pour les cliniciens et les patients. Il y a donc eu plusieurs approches à travers les années afin d'ajouter un apport psychométrique à l'évaluation clinique et au choix de traitement. Toutefois, il n'existe à ce jour aucune étude qui s'est penchée sur l'utilisation des innovations en santé numérique, telles que l'apprentissage machine, pour améliorer les approches psychothérapeutiques utilisées auprès des patients souffrant de schizophrénie afin de personnaliser le traitement prodigué. L'intégration de modalités issues de l'intelligence artificielle dans une psychothérapie spécialisée pour des patients atteints de schizophrénie avec des hallucinations auditives qui résistent à la médication nécessite un survol approfondi des thématiques suivantes : la schizophrénie, l'intelligence artificielle en psychiatrie et la réalité virtuelle. Ces trois éléments clés permettront aux lecteurs d'apprécier les objectifs principaux et les résultats présentés dans la section résultats de cette thèse.

La schizophrénie

Définition et évaluation clinique

Le trouble mental est une entité qui touche environ un Canadien sur trois dans son parcours de vie (Canadian Community Health Survey – Mental Health, 2012). Parmi les différentes psychopathologies appréciables dans le contexte de la psychiatrie moderne, la schizophrénie a une prévalence relativement faible (moins d'un pourcent de la population), mais représente un fardeau économique social significatif qui varient entre 94 à 102 milliards de dollars par année (Chong et al., 2017). Il s'agit d'une maladie ayant un impact négatif important pour les patients qui en souffrent puisqu'elle peut s'exprimer par une réduction de l'expression des émotions, une diminution de la motivation pour atteindre des objectifs personnels, des difficultés dans les relations sociales, une altération motrice et une détérioration des fonctions cognitives (Patel et al, 2014). Même si les facteurs de risques impliqués dans le développement de la schizophrénie chez un individu demeurent à l’étude, plusieurs pistes de réflexions sur le plan génétique, environnementaux et sociaux ont émergé dans les dernières décades (Pickard, 2011 ; Stilo et al., 2019).

Dans la littérature, il s’agit d’une maladie rapportée comme ayant une héritabilité élevée. Des facteurs génétiques et épigénétiques semblent être impliqués dans la pathogénèse de la schizophrénie avec des mécanismes tel que la méthylation de l'ADN et l'implication des éléments épigénétiques : 5-methycystosine et 5-hydroxylcytosine (Xie et al., 2024). De plus, dans l'étude de Hilker et al. (2018) qui utilise deux registres nationaux du Danemark afin d'évaluer le devenir de 31 524 paires de jumeaux, la prévalence de la schizophrénie sur le plan de l'héritabilité est estimée à 79%, ce qui supporte l'hypothèse de l'implication génétique dans la survenue de la maladie. Certains gènes impliqués dans le développement de syndromes génétiques, telle que la maladie cardio-vélo-faciale (délétion 22q11.2), l’allèle COMT et l’allèle PRODH s’inscriraient comme des sous-types de la schizophrénie (Bassett et al., 2008). Plusieurs autres hypothèses génétiques ont également été soulevées par le regroupement des études génomiques Genome-wide association studies (GWAS) tel que l'implication de polymorphismes nucléotidiques (SNP) pour six protéines impliquées dans le développement du système nerveux central (Dennison et al., 2019). Ce regroupement insiste sur le fait qu’il y aurait une panoplie de gène et de variantes génétiques impliqués et qui demeurent à être découverts. Parmi les facteurs de risque environnementaux liés à la schizophrénie, on peut citer, entre autres, le fait de vivre en milieu urbain durant l'enfance, l'immigration, un père plus âgé au moment de la naissance de l'individu, l'usage du cannabis, des expériences traumatisantes dans l'enfance, des infections chez la mère pendant la grossesse, ainsi que l'hypoxie autour de la naissance (Tandon et al., 2023).

Contemporainement, la schizophrénie se définie de plusieurs façons, mais les deux définitions les plus utilisées en clinique sont le celles issues du *Manuel diagnostique et statistique des troubles mentaux 5*$^{\text{ième}}$ édition (DSM-5) et de Classification Statistique Internationale des Maladies et des Problèmes de Santé Connexes 10$^{\text{ème}}$ édition (ICD-10). Cette thèse s’appuie sur la définition issue du DSM-5. Ce manuel défini la schizophrénie, en six critères, de la façon suivante (American

Psychiatric Association, 2022) :

A. Au moins deux des manifestations suivantes (une des manifestations doit être 1,2 ou 3) pendant au moins (ou moins si réponse favorable au traitement) :

    1. Idées délirantes

    2. Hallucinations

    3. Discours désorganisé

    4. Comportement grossièrement désorganisée ou catatonique

    5. Symptômes négatifs (par exemple un émoussement affectif, une alogie, une perte de volonté)

B. Durant une proportion significative de temps depuis le début du trouble, le niveau de fonctionnement dans un domaine majeur tel que le travail, les relations interpersonnelles ou l'hygiène personnelle est passé d'une façon marquée en dessous du niveau atteint avant le début du trouble (ou, quand le trouble apparaît pendant l'enfance ou l'adolescence, le niveau prévisible de fonctionnement interpersonnel, scolaire ou professionnel n'a pas été atteint).

C. Des signes continus du trouble persistent depuis au moins 6 mois. Pendant cette période de 6 mois les symptômes répondant au critère A (c'est à dire les symptômes de la phase active) doivent avoir été présents pendant au moins un mois (ou moins en cas de traitement efficace) ; dans le même laps de temps des symptômes prodromiques ou résiduels peuvent également se rencontrer. Pendant ces périodes prodromiques ou résiduelles, les signes du trouble peuvent ne se manifester que par des symptômes négatifs, ou par deux ou plus des symptômes listés dans le critère A présents sous une forme atténuée (par exemple des croyances étranges ou expériences de perceptions inhabituelles).

D. Le trouble schizo-affectif, ou dépressif, ou un trouble bipolaire, avec manifestations psychotiques ont été exclus parce que 1) soit il n'y a pas eu d'épisode maniaque ou dépressif caractérisé concurremment avec la phase la phase active des symptômes, 2) soit, si des épisodes de trouble de l'humeur ont été présents pendant la phase active des symptômes, ils étaient présents seulement pendant une courte période sur la durée totale des phases actives et résiduelles de la maladie.

E. Le trouble n'est pas imputable aux effets physiologiques d'une substance (p. ex. une drogue donnant lieu à abus, ou un médicament) ou à une autre pathologie médicale.

F. S'il existe des antécédents de trouble du spectre de l'autisme ou de trouble de la communication débutant dans l'enfance, le diagnostic surajouté de schizophrénie est posé seulement si des symptômes hallucinatoires et délirants importants, en plus des autres symptômes de schizophrénie nécessaires au diagnostic, sont aussi présents pendant au moins un mois (ou moins en cas de traitement efficace).

Sur le plan clinique, il est important de définir les éléments du critère A. Les idées délirantes représentent un ensemble de croyances fermes et erronées, lesquelles se confrontent directement à la réalité objective (López-Silva et al., 2024). Malgré la présence de preuves invalidantes, les individus atteints de délires demeurent incapables de renoncer à ces convictions erronées. Les idées délirantes sont donc souvent définies comme inébranlables (Kiran et al., 2009). Cette rigidité cognitive est un aspect critique de la pathologie délirante et permet de distinguer entre une simple croyance (ou pensée qu'un phénomène est vrai en l'absence de preuves tangibles). Les croyances délirantes sont souvent exacerbées par des interprétations inexactes ou biaisées des événements extérieurs, et elles s'accompagnent fréquemment de symptômes paranoïaques (Gipps et al., 2004). À titre illustratif, un individu en état de délire peut persister à croire, sans fondement probant, que des entités telles que l'armée exercent un contrôle omniscient sur les individus par des moyens technologiques avancés, tels que la télévision, l'ordinateur ou les ondes radio. Dans le spectre des troubles psychotiques, la présence de délires est un symptôme souvent prédominant (Rootes-Murdy et al., 2022). Ces derniers peuvent se manifester conjointement avec des hallucinations, qui se caractérisent par la perception de stimuli inexistants, comme avoir l'impression d'entendre des voix inaccessibles à autrui, de voir des choses qui n'y sont pas ou la sensation tactile de contact avec des objets imaginaires, tels que des insectes rampants sur la peau (Baker et al., 2019). Dans la schizophrénie, les délires les plus fréquents sont des délires de persécution (Stanghellini et al., 2015). C’est-à-dire que l’individu atteint de schizophrénie croit formellement qu’une entité lui veut du mal ou veut s’en prendre à autrui. Du côté des hallucinations, ce sont les hallucinations auditives qui prédominent (McLachlan et al., 2013). Les hallucinations et les délires sont également appelés symptômes positifs. Leurs origines sont complexes et toujours à l’étude. Toutefois, la littérature actuelle sur le sujet met l’emphase sur l’implication de deux neurotransmetteurs qui expliqueraient la survenue des symptômes positifs : la dopamine et le glutamate (Howes et al., 2015). Les études précliniques et pharmacologiques suggèrent l'implication de ces deux systèmes, et les observations in vivo révèlent une synthèse et une libération accrues de dopamine dans le striatum des schizophrènes. Une synthèse de la littérature récente démontre que les résultats concernant le système glutamatergique et certains aspects des études sur le système dopaminergique sont moins univoques, ce qui pourrait être attribué aux limites méthodologiques et à la diversité des manifestations de la schizophrénie (McCutcheon et al., 2020). Il en ressort que les facteurs génétiques et environnementaux associés à la schizophrénie influencent négativement les fonctions glutamatergiques et dopaminergiques. Toutefois, si les dysfonctions glutamatergiques semblent directement liées à la génétique, les anomalies du système dopaminergique seraient majoritairement dues à d'autres facteurs. Il y aurait également des interactions neuronales entre ces deux systèmes et leur déséquilibre pourrait engendrer les symptômes positifs.

Les symptômes négatifs sont présents chez environ 60 % des patients atteints de schizophrénie (An der Heiden et al., 2016). Ces symptômes se caractérisent par un affaiblissement ou une perte des fonctions et comportements habituels associés à l'envie et à l'expression, tant sur le plan verbal qu'émotionnel. Ils se manifestent à travers cinq aspects principaux : une émotion atténuée, l'alogie qui se traduit par une parole moins abondante, l'avolition marquant un déclin de l'activité orientée vers des objectifs spécifiques due à un manque de motivation, une tendance à l'asocialité, et l'anhédonie, qui est un appauvrissement de la capacité à ressentir du plaisir (Correll et al., 2020). Ces symptômes seraient secondaires à une hypoactivation des structures dopaminergiques du cortex mésocortical (Mosolov et al., 2022).

Sur le plan des déficits cognitifs, il a été largement établi que la schizophrénie est associée à des altérations dans plusieurs domaines cognitifs, notamment une diminution de l'efficacité de la mémoire de travail, de l'attention et de la vitesse de traitement de l'information (Tripathi et al., 2018). De plus, cette pathologie impacte l'apprentissage, tant visuel que verbal, et entraîne des déficits notables dans des fonctions cognitives supérieures telles que le raisonnement, la planification, la capacité de pensée abstraite et la résolution de problèmes (MacCabe et al., 2012). Ces déficits seraient présents chez la quasi-totalité des patients atteints de schizophrénie (Keefe et al., 2005).

Sur le plan épidémiologique, la schizophrénie touche autant les hommes que les femmes (Aleman et al., 2003). Toutefois, quelques études suggèrent que les hommes seraient davantage atteints (Nicole et al., 1992 ; Ochoa et al., 2012). Pour les hommes, l'apparition de la maladie suit un cours plutôt unimodal, se présentant habituellement entre l'âge de 18 et 25 ans, avec une proportion de symptômes négatifs plus élevée que chez la contrepartie féminine (Li et al., 2016). Il y aurait également une durée plus longue de la maladie et un moins bon pronostic. Du côté des femmes, la maladie suit un courant bimodal, avec deux pics d'apparition : 25-35 ans et un deuxième pic après 40 ans (Li et al., 2022). Il y aurait davantage de symptômes affectifs et de symptômes positifs (Thara et al., 2015).

Avenues thérapeutiques

Le traitement de la schizophrénie est multimodal. Celui-ci repose principalement sur des approches pharmacologiques, considérant les implications neurophysiologiques de la maladie, et sur des approches psychothérapeutiques visant à aider le patient dans la gestion du fonctionnement altéré et la maîtrise de sa maladie. D'autres approches ponctuelles, telles que l'électroconvulsivothérapie, la neuromodulation et les approches systémiques présentent des niveaux de preuve variés et ne seront pas abordées dans le cadre de cette thèse (Kirkpatrick et al., 2014 ; Dokucu, 2015).

Avenues pharmacologiques

Dans le cadre du traitement pharmacologique de la schizophrénie, l'accent est principalement mis sur les symptômes positifs, en raison de leur lien avec des conséquences parfois sévères, telles que le suicide ou la violence. Il est observé que la diminution de ces symptômes peut entraîner une amélioration des symptômes négatifs (Tandon, 2011). Les auteurs d'une méta-analyse récente ont examiné les recommandations et algorithmes de consensus à l'échelle internationale concernant le traitement pharmacologique de la schizophrénie, identifiant les 19 principaux guides de recommandations (Correll et al., 2022). Toutefois, il existe un manque d'études sur l'application concrète de ces recommandations dans la pratique clinique quotidienne. Ce qu'il en ressort, c'est qu'à l'heure actuelle, les médicaments antipsychotiques demeurent la modalité de choix dans le traitement pharmacologique de la schizophrénie (Remington et al., 2017).

En lien avec les implications de la dopamine dans l'apparition des symptômes positifs de la schizophrénie, les antipsychotiques sont des médicaments qui, par définition, bloquent les récepteurs de la dopamine (récepteurs D2). L'élaboration de ces médicaments dans les années 1950 a constitué le point de départ de l'ère contemporaine du traitement pharmacologique de la schizophrénie. Les premiers antipsychotiques, qui antagonisent de façon importante les récepteurs D2 tout en ayant un effet moindre sur d'autres types de récepteurs, sont connus sous le nom de la grande famille des antipsychotiques typiques. En bloquant les récepteurs dopaminergiques D2 au niveau du système mésolimbique, ces médicaments permettent l'atténuation des symptômes positifs. Toutefois, un problème majeur avec ces molécules est qu'elles affectent non seulement les récepteurs dopaminergiques D2 du système mésolimbique, mais également l'ensemble des voies dopaminergiques cérébrales, entraînant des effets secondaires tels que des troubles du mouvement ou des effets liés à la régulation de la prolactine (par exemple, la galactorrhée ou la gynécomastie) (Stroup et al., 2018). Celles-ci comprennent principalement :

1. La voie mésolimbique : La voie dopaminergique mésolimbique représente une trajectoire

essentielle dans la modulation des mécanismes de récompense, de motivation et de gestion des émotions dans le cerveau des humains. Son origine se situe au niveau de l'aire tegmentale ventrale, localisée dans le mésencéphale, et ses projections s'étendent principalement vers le noyau accumbens, qui est également situé dans le système limbique du cerveau (McCutcheon et al., 2019). Cette voie cible également d'autres structures telles que l'amygdale (aussi connu comme étant le centre de la peur), l'hippocampe (impliqué dans la mémoire et les apprentissages) et certaines régions du cortex préfrontal (Richter-Levin, 2004). Sa fonction primordiale réside dans son rôle dans le renforcement positif et la motivation, réagissant activement aux stimuli gratifiants et participant à la genèse des sensations de plaisir (Estave et al., 2022). Elle influence également divers comportements motivés, tels que la quête de nourriture, d'activité sexuelle, d'interactions sociales, et module la réponse aux substances addictives (Estave et al., 2022). Dans la la schizophrénie, le système mésolimbique serait impliqué dans les symptômes positifs. Selon l'hypothèse prédominante actuelle pour expliquer les symptômes positifs, une suractivité dopaminergique dans le noyau accumbens ainsi que dans d'autres composantes de ce système serait responsable de ces manifestations (Richter et al., 2015). Cette hyperactivité pourrait être le résultat d'une dysrégulation dans la libération de dopamine ou d'une hypersensibilité des récepteurs dopaminergiques au sein de ces zones cérébrales (Richter et al., 2015).

2. La voie mésocorticale : La voie dopaminergique mésocorticale, joue un rôle dans la régulation des fonctions cognitives et de la régulation des émotions. Elle origine également à partir de l'aire tegmentale ventrale et se projette vers le cortex préfrontal (McCutcheon et al., 2019). Ce cortex contrôle de la cognition, la prise de décision et la gestion des émotions. Il intervient également de manière significative dans des domaines tels que la concentration, la planification stratégique, la résolution de problèmes complexes et la gestion des émotions (Friedman, 2022). Sa fonction s'étend également à la réflexion abstraite, au jugement et à la facilitation des interactions sociales (Friedman, 2022). Dans le cadre spécifique de la schizophrénie, les anomalies fonctionnelles au sein de ce système dopaminergique sont fortement corrélées aux symptômes négatifs et aux

troubles cognitifs observés dans cette pathologie (Brisch et al., 2014). Les troubles cognitifs associés peuvent se manifester par des difficultés liées à la mémoire de travail, à la concentration et à d'autres aspects des fonctions exécutives (Robison et al., 2020). L'origine de ces symptômes serait attribuée à un déficit de l'activité dopaminergique au sein du cortex préfrontal.

3. La voie nigro-striée : La voie dopaminergique nigro-striée est associée à la régulation des fonctions motrices. Elle se situe au niveau de la substance noire et se projette vers le striatum (noyau caudé et le putamen) (McCutcheon et al., 2019). Son rôle est principalement de coordonnée les mouvements volontaires, effectuée le nécessaire pour permettre l'exécution des activités motrices et est impliquée dans l'équilibre postural incluant la motricité fine. Un blocage de l'activité dopaminergique de ces voies est donc impliqué avec des troubles du mouvements (Glazer, 2000).

4. La voie tubéro-infundibulaire : À l'instar des autres voies, la voie dopaminergique tubéro-infundibulaire se situe dans la région tubérale de l'hypothalamus et se projette vers l'infundibulum et l'hypophyse (McCutcheon et al., 2019). Elle est impliquée dans la régulation de la sécrétion de la prolactine (Grattan, 2015). Dans la schizophrénie, les antipsychotiques (étant donné leur antagonisme sur le plan des récepteurs dopaminergiques) peuvent permettre un relâchement accru de prolactine amenant à des effets secondaires indésirables conséquentes à cette augmentation de prolactine (Goodnick et al., 2002).

À cet effet, une deuxième génération d'antipsychotiques a vu le jour, aussi connue sous le nom d'atypiques. Celle-ci se caractérise par un blocage sérotoninergique au niveau du récepteur 5-HT2A tout en maintenant l'antagonisme du récepteur dopaminergique D2 (Seeman, 2002). Les médicaments tels que l'olanzapine, la rispéridone, la lurasidone et la clozapine sont des exemples de cette deuxième génération. Le principal avantage de ces molécules est la diminution des effets secondaires liés au blocage dopaminergique important de la première génération d'antipsychotiques (Wright et al., 2003). Toutefois, en raison de leurs activités variées au niveau des récepteurs histaminiques, alpha-adrénergiques, muscariniques et autres, les antipsychotiques de deuxième génération ont des impacts métaboliques importants, entraînant comme principaux effets secondaires une prise de poids, une augmentation de la prévalence du diabète, des syndromes cardiovasculaires, de l’hypertension et de la dyslipidémie (Toren et al., 2004).

Un troisième type d’antipsychotique est apparu, les agonistes partiels des récepteurs dopaminergiques D2, tels que l’aripiprazole et le brexpiprazole (Kishi et al., 2020). Ces molécules sont également utilisées dans la schizophrénie et ont comme avantage observé de réduire la survenue des effets secondaires métaboliques observés avec les autres antipsychotiques de deuxième génération (Stip et al., 2010).

À ce jour, il n'existe pas de médicament permettant de traiter les symptômes de la schizophrénie chez tous les patients atteints (Leucht et al., 2013). Les profils de réponse et de tolérance aux antipsychotiques sont variés, ce qui nécessite des approches personnalisées pour les patients vivant avec les conséquences de la schizophrénie (Leucht et al., 2013 ; Lally et al., 2015).

Avenues psychothérapeutiques

L'approche psychothérapeutique ayant le plus de preuves cliniques actuellement pour les maladies impliquant une psychose (critère A de la schizophrénie) est la thérapie cognitivo-comportementale pour la psychose (TCCp) (Kart et al., 2021). De façon générale, la thérapie cognitivo-comportementale (TCC) est une méthode d'intervention psychologique conçue pour aider les personnes aux prises avec des distorsions cognitives importantes en remettant en question leur modèle de pensée (Wenzel, 2017). Elle est employée depuis les années 1970 pour traiter divers types de psychopathologies et gérer les troubles émotionnels (Hofmann et al., 2012). Cette approche intègre diverses techniques, incluant l’exploration des croyances, des éléments de thérapie comportementale, des stratégies de résolution de problèmes et l'entraînement à des techniques de gestion du stress (Nakao et al., 2021). La TCCp s'est révélée avoir des effets positifs sur les symptômes positifs et négatifs de la schizophrénie (Turner et al., 2014). Elle est couramment utilisée en tant que thérapie complémentaire aux médicaments, conformément à certains guides de pratique (Avasthi et al., 2020). Toutefois, d'après les conclusions récentes de plusieurs analyses Cochrane, la qualité insuffisante des données actuellement disponibles ne permet pas d'établir des conclusions définitives sur l'efficacité de la TCC en tant qu'ajout au traitement habituel pour les personnes souffrant de schizophrénie, ni sur son efficacité comparée à d'autres interventions psychosociales (Ballesteros et al., 2023). Au Canada, une évaluation récente en Ontario suggère que l’ajout de la TCCp au traitement habituel permet toutefois d’obtenir davantage de bénéfices en termes de réduction des hallucinations auditives et des délires chez les patients atteints de psychose au sens large (Health Quality Ontario, 2018). Concernant la qualité de vie, les auteurs d'une analyse de 36 essais cliniques randomisés ont montré de très faibles bénéfices en termes de réduction de la détresse liée aux symptômes négatifs et aucune amélioration sur le plan de la qualité de vie (Laws et al., 2018). Comme la schizophrénie se présente sur un spectre avec une variété de symptômes positifs et négatifs, plusieurs formes de TCCp ont été développées, ciblant différents aspects (Rathod et al., 2020). Les principes de base sous-tendent que le stress, au cœur des manifestations liées aux symptômes positifs, fait émerger des pensées ancrées dans des croyances fondamentales difficilement ébranlables (Abdel-Baki et al., 2001). En ce sens, une pensée négative serait liée à des émotions négatives et à l’adoption de comportements délétères pour le patient. La TCCp vise donc à modifier les pensées pour influencer les émotions et les comportements qui y sont associés. Elle peut être donnée sous forme individuelle ou en groupe. Considérant la variété des résultats des études portant sur la TCCp pour les patients atteints de schizophrénie et de schizophrénie réfractaire à la médication, l’utilisation de la TCCp reste une question d’actualité, malgré son inclusion dans plusieurs guides de pratique (Morrisson, 2009).

Dans le traitement de la schizophrénie, plusieurs types de thérapies issues des courants relationnels et humanistes ont émergé au fil des années (Shattock et al., 2018). Les courants relationnels considèrent davantage les symptômes positifs de la schizophrénie comme un mode d’interaction entre le patient et son environnement (Thomas et al., 2014). Dans ce contexte, les hallucinations auditives sont par exemple interprétées comme une entité avec laquelle le patient aux prises avec la schizophrénie doit interagir au quotidien (Hayward et al., 2009). Certains patients personnifient donc les voix qu’ils entendent et c’est dans cette optique que la thérapie s’inscrit. Dans ce cadre, le patient peut mettre en pratique les dynamiques interpersonnelles qu’il vit et subit à travers ses interactions avec les hallucinations auditives (Alderson-Day et al., 2020). L'analyse de 17 études sur le sujet par Dellazzizo et al. (2022) témoigne de l'existence de plusieurs types de thérapies issues du courant relationnel, mais indique que majoritairement, la TCC orientée vers la thérapie relationnelle et la TA seraient actuellement les plus utilisées. La TA sera discutée en détail dans la section portant sur la réalité virtuelle.

La schizophrénie réfractaire à la médication

Malgré la panoplie de traitements disponibles, il est reconnu qu'un pourcentage important de patients atteints de schizophrénie résiste au traitement, avec une persistance de leurs symptômes positifs. Une revue de la littérature, dans laquelle les auteurs s’intéressent aux données de 29 390 patients atteints de schizophrénie, estime qu'il y a une prévalence de résistance à la médication d'environ 23,6 % (Diniz et al., 2023). D'autres études observationnelles estiment que cette prévalence se situe entre 20 et 50 % (Beck et al., 2019). L'enjeu persiste toutefois à définir ce qu'est la schizophrénie réfractaire à la médication (SRT) et comment elle se définit.

La convention la plus acceptée concernant la schizophrénie résistante au traitement (SRT) est celle où un patient, aux prises avec des symptômes positifs, continue de présenter ces symptômes malgré au moins deux essais de médicaments antipsychotiques à des doses et durées adéquates, avec une adhésion documentée (Nucifora et al., 2019).

La première ligne de traitement pour ces patients est le passage à la clozapine, qui est un antipsychotique de deuxième génération. Cependant, il est estimé que 40 à 70 % des patients sous clozapine deviendront éventuellement résistants à cette molécule et qu’il sera nécessaire de se tourner vers d'autres formes de traitements (Chakrabarti, 2021). La littérature actuelle démontre des effets variés quant aux stratégies d'augmentation avec d'autres médicaments, à l'électroconvulsivothérapie ou à la neuromodulation pour les patients qui présentent des symptômes positifs malgré une prise adéquate de clozapine (Chiu et al., 2020). Il y a donc une nécessité d'explorer davantage d'alternatives thérapeutiques pour ces patients.

Trajectoires cliniques

Au fil des années, de nombreux efforts de recherche ont été déployés afin de prédire quels patients atteints de schizophrénie auront une bonne évolution clinique et quels autres connaîtront une évolution défavorable en termes de multiples facteurs tels que la symptomatologie, la qualité de vie et le développement personnel et professionnel (Zipursky, 2014). Les études sur le sujet révèlent une difficulté inhérente à cet exercice de classification en raison de l'hétérogénéité de la maladie et des personnes qui en souffrent. La revue systématique et la méta-analyse de Molstrom et al. (2022), portant sur 14 études prospectives (1 991 patients), ont rapporté qu'un rétablissement s'est manifesté chez 24,2 % des patients, 35,5 % des patients avaient de bons résultats cliniques en termes de symptomatologie, et le reste a connu des améliorations plutôt modérées. Toutefois, ces données concernent des patients ayant reçu un traitement et suivis par des équipes médicales, ce qui n'est pas toujours le cas.

À cet effet, les facteurs de bonne évolution clinique comprendraient : un début soudain de la maladie, un âge d'apparition plus tardif, l'absence de schizophrénie ou de trouble psychotique dans la famille, des symptômes positifs (plutôt que négatifs), une bonne autocritique par rapport à la maladie, la présence de symptômes affectifs associés, une bonne observance du traitement, un bon fonctionnement pré-morbide et inter-épisode psychotique, un fonctionnement neurologique normal et un bon réseau de soutien (Kay et al., 1990 ; Emsley et al., 2008 ; Schennach-Wolff et al., 2009 ; Peña et al., 2012 ; van Dee et al., 2023).

Il s'agit néanmoins d'une psychopathologie très complexe et les études ne s'accordent pas toutes sur ces prédicteurs. De manière générale, il est estimé que dans la trajectoire d'un patient atteint de schizophrénie, environ un tiers connaît une bonne évolution avec une rémission et les deux tiers restants présentent des symptômes intermittents sur l'évolution chronique de la maladie (Liebermann, 2006 ; George et al., 2017).

De plus, l'utilisation de cannabis a récemment été identifié comme un facteurs prédisposant à la psychose et devançant l'apparition de trouble psychotique primaire pour les groupe de patients à risque (Starzer et al., 2018). Une revue de littérature sur le sujet regroupant 12 articles, traitant de l'utilisation du cannabis, de la violence et des patients atteints de schizophrénie, rapporte une association modérée entre l'utilisation de cannabis et la violence pour ces patients (Dellazizzo et al., 2019).

Conclusion de la section

En conclusion, la schizophrénie est une pathologie complexe pour laquelle les traitements actuels permettent partiellement le rétablissement des patients qui en sont atteints. L'approche favorisée pour le traitement de la schizophrénie est une approche multimodale, impliquant des antipsychotiques pour cibler les symptômes positifs de la maladie. Vu le grand nombre de patients qui ont des symptômes qui résistent à la médication, des stratégies psychothérapeutiques et autres stratégies d'appoint sont utilisés. C'est dans ce contexte que s'inscrit la TA. Considérant l'hétérogénéité de la maladie et de l'évolution de celle-ci, il apparaît important de personnaliser davantage le traitement pour les personnes atteints de schizophrénie afin de favoriser une amélioration clinique sur le plan de la symptomatologie et de la qualité de vie.

L'intelligence artificielle en psychiatrie

Face à l'augmentation de la complexité des patients atteints de psychopathologies rencontrés dans la pratique clinique des psychiatres et autres professionnels de la santé, la psychiatrie computationnelle a émergé comme un outil prometteur (Montague et al., 2012 ; Wang et al., 2014). Cette discipline, issue des sciences informatiques, se définit par l'application de modèles mathématiques, tels que les statistiques, aux cadres théoriques pour aborder et résoudre des problématiques en psychiatrie clinique (Adams et al., 2016; Maia et al., 2017). Un domaine spécifique de la psychiatrie computationnelle implique l'utilisation de l'intelligence artificielle, offrant ainsi de nouvelles perspectives et approches dans l'étude et le traitement des troubles neuropsychiatriques (Jin et al., 2023).

Définitions

L'apprentissage machine est une branche spécifique de l'intelligence artificielle. Les modèles d'apprentissage machine sont des algorithmes statistiques conçus pour identifier et apprendre les attributs significatifs (*features*) des données et leur importance relative, essentiels pour la prédiction d'une variable ciblée, ou pour évaluer l'importance des caractéristiques des données définies par l'utilisateur. Lorsque les attributs, leurs poids, ainsi que d'autres paramètres sont déterminés, le modèle est capable de réaliser des prédictions en lien avec une issue spécifique ou d'effectuer une classification clinique basée sur de nouvelles données encore inexplorées. Il existe principalement deux types d'apprentissage machine : supervisé et non-supervisé. Dans l'apprentissage machine supervisé, les données avec lesquelles le modèle est entraîné ont des libellés : elles sont classifiées au-préalable dans des catégories distinctes. Au contraire, l'apprentissage machine non-supervisé utilise les données brutes et générera des catégories en lien avec paramètres (par exemple des ressemblances sur le plan des données).

Applications cliniques en psychiatrie

Les applications de l'intelligence artificielle en psychiatrie, à l'heure actuelle, se situent principalement autour de la prédiction, du diagnostic et du traitement de maladie psychiatrie et des patients qui en sont atteints (Fakhoury, 2019). Plus récemment, il y a émergence d'outils de qualité hétérogènes axés sur le suivi des symptômes, sur la psychoéducation et l'aide au patient par robots clavardage augmenté par intelligence artificielle (Pham et al., 2022). Les algorithmes les plus fréquemment utilisés sont issues des techniques d'apprentissage machine et du traitement du langage naturel (Chandler et al., 2020).

L'utilisation de l'apprentissage machine et ses applications

L'apprentissage machine supervisé, utilisant des données préalablement catégorisées, permet d'entraîner des algorithmes (de classification ou de prédiction) à classer de nouvelles données selon des paramètres définis ou à prédire des résultats spécifiques (Jiang et al., 2020). Cette approche englobe divers algorithmes, basés sur des fondements statistiques et ayant des utilités distinctes, pour la classification des données, incluant les classificateurs linéaires, les machines à vecteurs de support et les arbres de décision (Sarker, 2021). En parallèle, il existe de nombreux algorithmes prédictifs tels que la régression linéaire, la régression logistique et la régression polynomiale, pour n'en citer que quelques-uns (Nichols et al., 2019). Face à la complexité de divers tableaux cliniques en psychiatrie, l'apprentissage machine supervisé offre aux cliniciens des outils puissants pour exploiter une variété de données. Cela pourrait accélérer la détection de certaines psychopathologies, proposer des traitements personnalisés et limiter les conséquences négatives de l'évolution d'une maladie psychiatrique. De nombreux auteurs s'accordent sur l'intérêt d'intégrer des marqueurs biologiques, signes et symptômes, imageries et autres sources de données pour affiner et personnaliser les traitements, comme cela est observé dans d'autres domaines de la médecine (Cearns et al., 2019). Par exemple, pour les maladies psychiatriques ayant une forte composante génétique, il serait envisageable d'utiliser un algorithme de classification pour identifier le profil génétique d'un patient et vérifier s'il correspond à des phénotypes psychiatriques spécifiques, afin d'initier des traitements préventifs ciblés. En 2019, une revue de littérature analysant 300 articles a identifié quatre domaines d'applications cliniques pertinents de l'apprentissage machine : (1) détection et diagnostic ; (2) évolution, traitement et accompagnement du patient ; (3) santé de la population ; (4) recherche et gestion des soins de santé (Shatte et al., 2019).

L'apprentissage machine non-supervisé, quant à lui, analyse des données non prédéfinies ou non classées (Usama et al., 2019). La pratique clinique en psychiatrie, en termes de diagnostic, repose sur des paradigmes catégoriels tels que le DSM-5 ou dimensionnels (Potusak et al., 2012). Devant l'hétérogénéité des symptômes psychiatriques, l'approche catégorielle est souvent débattue dans la communauté scientifique, car elle ne parvient pas toujours à cerner précisément certaines psychopathologies et leurs évolutions. Par exemple, chez les patients atteints de schizophrénie, certains répondent au traitement tandis que d'autres non : appartiennent-ils vraiment à la même catégorie diagnostique ou existent-il des sous-groupes avec des phénotypes cliniques distincts ? L'apprentissage machine non-supervisé aide à explorer ces questions en identifiant, à partir de grandes quantités de données, des regroupements de données similaires. Cela permet d'aborder les sous-types des maladies psychiatriques de façon trans-diagnostique, en fonction de caractéristiques spécifiques (Pelin et al., 2021). Cette méthode facilite également l'intégration de données multidimensionnelles dans la compréhension clinique de certaines psychopathologies, comme l'association des critères diagnostiques d'un trouble de l'attention avec des données génétiques (Zhao et al., 2018). En pratique clinique, cela permet au clinicien de mieux appréhender les particularités et l'intensité des symptômes et de la souffrance clinique d'un patient, et d'ajuster le plan de traitement en conséquence. Par exemple, une étude chinoise portant sur plus de 8 000 patients a démontré l'efficacité d'une approche non supervisée pour identifier avec plus de précision l'intensité d'un trouble dépressif par rapport aux outils de dépistage traditionnels (Yang et al., 2020). Une autre étude, plus récente, utilise l’apprentissage machine non-supervisé pour interpréter les données issues de montres intelligentes pour différencier les patients atteints de schizophrénie avec des symptômes affectifs de patients atteints de dépression avec plus de précision en se penchant sur les changements comportementaux (Price et al., 2022).

Le traitement du langage naturel

De façon générale, le traitement du langage naturel, un domaine de l'intelligence artificielle qui modélise le langage humain, a été utilisé en médecine pour automatiser les diagnostics, détecter par exemple les événements indésirables des traitements, soutenir la prise de décision clinique et prédire les résultats cliniques (Doan et al., 2014). En psychiatrie, le traitement du langage naturel a trouvé des applications spécifiques, telles que la synthèse automatique de la littérature, l'extraction de données pertinentes, l'identification des patients, la rédaction automatique de rapports cliniques et la prédiction des issues thérapeutiques (Le Glaz et al., 2021).

Cette technologie permet non seulement de gérer efficacement d'immenses volumes de données textuelles, mais aussi d'interpréter le langage des patients, souvent riche en nuances, intensité et en subtilités, pour en extraire des informations cliniquement pertinentes (Khanbhai et al., 2021). Ainsi, en psychiatrie, le traitement du langage naturel peut contribuer par exemple à affiner les diagnostics, à identifier des patterns dans les symptômes rapportés par les patients, à analyser les notes cliniques pour une meilleure compréhension des trajectoires de soins et à prédire les réponses aux traitements (Crema et al., 2022). Cet outil peut également aider les cliniciens et les chercheurs à rester à jour avec les avancées rapides dans le domaine, en synthétisant de grandes quantités de publications et de rapports de recherche (Scaccia et al., 2021). Cette approche s’inscrit donc favorablement dans la quête d'une médecine psychiatrique plus précise et personnalisée.

Médecine personnalisée en psychiatrie

Face à la diversité des réactions des individus atteints de troubles mentaux aux traitements existants, la pensée, remontant même à l'époque d'Hippocrate, suggère que les êtres humains, de par leur complexité et leurs particularités physiologiques, présentent une susceptibilité unique à certaines pathologies et à leurs réponses thérapeutiques (Ozomaro et al., 2013). Cette perspective a conduit à une évolution dans les pratiques cliniques, reconnaissant la nécessité d'approches de soins personnalisées pour chaque patient (Quinlan et al., 2020). Ainsi, la psychiatrie clinique s'oriente progressivement vers la médecine de précision, un volet de la médecine personnalisée, visant à exploiter des indicateurs quantifiables relatifs aux patients, à leurs symptômes, diagnostics, pronostics et réponses aux traitements (Češková et al., 2021).

Au cours de la dernière décennie, la médecine de précision a de plus en plus recours à des modèles mathématiques issus de l'intelligence artificielle, notamment les algorithmes d'apprentissage machine (Manchia et al., 2020). Ces approches pluralistes intègrent une multitude de variables, telles que des indicateurs biologiques, des données d'imagerie, des informations extraites des dossiers médicaux, des mesures biométriques et autres, dans l'élaboration de ces modèles (Tonelli et al., 2017).

Du point de vue technique, la médecine de précision marque un progrès notable en termes de spécificité des résultats étudiés (Roche et al., 2021). Cependant, elle est soumise à de nombreuses contraintes techniques, car les modèles mathématiques employés dépendent fortement de la qualité, de la quantité des données disponibles, du contexte culturel et social de leur collecte, ainsi que des indicateurs de performance utilisés pour évaluer leur validité (Naithani et al., 2021).

Enjeux éthiques et défis

Plusieurs enjeux éthiques et défis d’utilisation de l’intelligence artificielle dans le domaine de la psychiatrie clinique existent et doivent être pris en considération. Selon l'étude de Koutsouleris et al. (2022), l'implémentation de l'IA en psychiatrie se heurte à divers obstacles, notamment en ce qui concerne le choix des modèles mathématiques, la sélection des paradigmes pour les résultats cliniques envisagés, la complexité biopsychosociale des troubles mentaux, la quantité de données nécessaires pour former les modèles mathématiques et la question de la validation externe.

D'un point de vue éthique, de multiples interrogations émergent, notamment sur l'application des algorithmes à de nouveaux patients (touchant ainsi à la validité externe), la transparence des modèles utilisés et la préservation de la confidentialité des données. Par exemple, dans le domaine de la schizophrénie, Chekroud et al. (2024) ont observé que les modèles d'apprentissage machine, bien qu'intégrant des données issues de cinq essais cliniques randomisés, n'ont pas réussi à prédire efficacement la réponse au traitement chez les patients atteints de schizophrènie. Cette limitation pourrait être attribuée à la diversité des contextes des bases de données, à l'insuffisance des données initiales pour l'entraînement des algorithmes et à la nature potentiellement contextuelle des issues de la schizophrénie (Chekroud et al., 2024).

D’autres défis tel que le surajustement (overfitting) ou le sous-ajustement (underfitting) peuvent survenir dans la conception des modèles prédictifs. Le surajustement se produit lorsqu'un modèle d'apprentissage automatique devient trop complexe et apprend à la fois les données d’entraînement et le bruit dans celles-ci (Demšar et al., 2021). Cela conduit à une grande performance lorsque testé sur les données d'entraînement, mais à de mauvaises performances sur des données de test, car le modèle ne généralise pas bien. Cela affecte donc directement la validité externe (Demšar et al., 2021). Ce problème est souvent causée par des modèles trop complexes, un manque de données d'entraînement ou l'absence de régularisation (Demšar et al., 2021).. Pour atténuer le surajustement, des stratégies telles que la simplification du modèle, l'augmentation de la taille des données, l'utilisation de la validation croisée et l'application de techniques de régularisation sont employées (Tufail et al., 2023). De l'autre côté, le sous-ajustement se produit lorsqu'un modèle est trop simple pour capturer les complexités des données d'entraînement, entraînant de mauvaises performances à la fois sur les données d'entraînement et de test (Belkin et al., 2019). Cela est généralement dû à des modèles trop simplistes, ou à une régularisation excessive des variables utilisées dans le modèle (Belkin et al., 2019). Pour aborder le sous-ajustement, il faut augmenter la complexité du modèle, améliorer le choix des variables, réduire la régularisation et éventuellement choisir un modèle plus adapté au contexte de la tâche à effectuer par le modèle (Joodaki et al., 2021).

Les études initiales concernant l'acceptabilité de l'IA comme outil d'amélioration des soins cliniques indiquent une réception positive de la part des patients, des cliniciens et du grand public (Young et al., 2021 ; Kleine et al., 2023). Néanmoins, les questions éthiques restent prépondérantes et figurent parmi les principales préoccupations des patients et de la population générale (Young et al., 2021).

Conclusion de la section

L'intelligence artificielle, appliquée en psychiatrie, connaît un développement notable. Elle suscite des attentes prometteuses en matière de médecine personnalisée, en s'appuyant sur l'utilisation de modèles mathématiques. Pour les patients souffrant de SRT, ces approches pourraient s'avérer bénéfiques pour une compréhension approfondie de la pathologie et des cibles cliniques spécifiques. Par exemple, le recours à l'apprentissage machine a révélé des facteurs prédictifs de la qualité de vie en exploitant les données issues de la vaste étude CATIE2, incluant trois groupes de patients atteints de schizophrénie (Beaudoin et al., 2022). Cependant, il convient d'aborder avec prudence l'interprétation des résultats et l'usage de modèles mathématiques, en considérant les multiples implications éthiques, mathématiques et les défis pour garantir leur pertinence dans le contexte clinique. À cet égard, exposer clairement les limites des modèles, assurer une transparence concernant les données utilisées et le type de modèles mathématiques sélectionnés pourrait contribuer à une meilleure compréhension de l'utilité clinique spécifique des modèles développés (Young et al., 2021).

La réalité virtuelle

Définitions

Depuis ses débuts dans les années 1950 et son intégration croissante dans les pratiques thérapeutiques innovantes pour les troubles psychiatriques dès les années 90, la RV a été annoncée comme un outil prometteur dans la personnalisation du traitement des patients souffrant de psychopathologies complexes (Park et al., 2019). Cette technologie se distingue par trois attributs essentiels : son caractère immersif, son interactivité, et sa capacité à fournir un retour sensoriel (Dellazizzo et al., 2020 ; Albakri et al., 2022).

De façon plus large, il existe des classifications spécifiques pour décrire ce qu’est la RV. Une classification initiale utilisée dans le domaine de la RV, basée sur la taxonomie de Vergara a été établie en 2017 (Vergara et al., 2017). Cette taxonomie distingue deux types de réalité virtuelle : non-immersive et immersive. Le type non-immersif se caractérise par l'emploi de dispositifs numériques standards, tels que des ordinateurs, et peut inclure l'utilisation d'écrans de grande taille. En contraste, la réalité virtuelle immersive englobe des environnements comme les voûtes virtuelles et l'utilisation de casques haptiques. Les caves virtuelles, grâce à des écrans stéréoscopiques et des technologies de suivi de mouvement de la tête, créent une expérience d'immersion et de présence (Moscoso et al., 2022). L'utilisateur peut ainsi interagir de façon dynamique avec l'environnement virtuel. Par ailleurs, un casque haptique de réalité virtuelle offre une expérience immersive grâce à une combinaison de stimuli visuels et, parfois, auditifs, procurant une impression de réalité tridimensionnelle (Kern et al., 2020 ; Turso-Finnich et al., 2023).

Une autre perspective sur la réalité virtuelle et ses variantes émane du concept de réalité mixte (Kolecki et al., 2022). Introduit par Milgram et Kishino en 1994, ce concept établit un continuum entre l'environnement réel et la réalité virtuelle (Skarbez et al., 2021). Ce continuum comprend quatre éléments : l'environnement réel, la réalité augmentée, la virtualité augmentée et, enfin, la réalité virtuelle (Milgram & Kishino, 1994). La réalité augmentée fusionne des éléments numériques avec le monde réel, permettant aux utilisateurs d'interagir avec des contenus virtuels intégrés dans leur environnement réel (Cipresso et al., 2018). Cette expérience peut combiner des éléments réels et virtuels, en superposant par exemple des objets réels sur un arrière-plan virtuel ou en intégrant des données en temps réel dans un environnement virtuel. Enfin, dans ce paradigme, la définition de la réalité virtuelle correspond à celle de la réalité virtuelle immersive définie par la taxonomie de Vergara (Vergara et al., 2017).

L’utilisation de la réalité virtuelle en psychiatrie

Dans le domaine de la psychiatrie clinique, une étude récente de Cieślik et al. (2020) a classifié les applications thérapeutiques de la RV en quatre catégories principales : les phobies et troubles anxieux, la gestion de la douleur, les troubles neurodéveloppementaux et autres troubles mentaux. Parallèlement, Dellazizzo et al. (2020), ont également répertorié à travers 11 méta-analyses les domaines d’utilisation de la RV à travers les pathologies suivantes : troubles anxieux, phobies spécifiques, anxiété sociale, traumatismes, troubles neurodéveloppementaux, troubles neurocognitifs et les troubles mentaux sévères (trouble dépressif, schizophrénie). La qualité des études recensées à travers ces revues de la littérature témoigne de qualité de preuves plutôt faible à modéré quant à l’efficacité de ces thérapies pour ces troubles mentaux.

La thérapie par Avatar

La thérapie par avatar est une avancée majeure dans la prise en charge de la schizophrénie. Elle s'avère particulièrement efficace pour les patients souffrant d'hallucinations auditives résistantes aux traitements pharmacologiques. Cette méthode se distingue par son caractère innovant et sa capacité à s'adapter aux besoins individuels de chaque patient, offrant une alternative significative aux méthodes thérapeutiques conventionnelles. L'utilisation de la réalité virtuelle pour créer un avatar numérique représentant l'hallucination auditive du patient est incluse dans cette méthode, qui a été développée par Leff et al. (2013) en 2008. L'objectif principal de cette thérapie est de matérialiser l'hallucination afin que le patient puisse l'affronter directement et développer un contrôle dans le contexte de ses interactions avec celle-ci. Les recherches, notamment celles menées par Craig et al. (2018) et Dumais et al., ont démontré que cette méthode est efficace pour réduire la fréquence et l'intensité des hallucinations auditives (Dellazizzo et al., 2021). Les patients peuvent atténuer leur souffrance psychologique en apprenant des stratégies de gestion de leurs symptômes en interaction avec l'avatar.

Parmi les grandes études sur le sujet, nous retrouvons tout d'abord l'étude de Craig et al. (2018). Les auteurs de l'étude ont mené un essai contrôlé randomisé en simple aveugle pour évaluer la TA, en la comparant aux approches de soutien dans le traitement des hallucinations auditives verbales chez les individus atteints de troubles du spectre de la schizophrénie ou de troubles affectifs. Les participants ont été assignés au hasard soit à la TA, soit aux approches de soutien, et évalués à l'aide de l'échelle de notation des symptômes psychotiques - hallucinations auditives (PSYRATS–HA) au départ, à 12 semaines et à 24 semaines. Les résultats ont indiqué une réduction statistiquement significative des scores PSYRATS–HA avec la TA par rapport aux approches de soutien à 12 semaines, avec une différence moyenne de –3,82 et une taille d'effet (d de Cohen) de 0,8, sans aucun événement indésirable attribuable à l'une ou l'autre thérapie (Craig et al., 2018). Ces résultats sont comparables à ceux observés dans du Sert et al. (2018). Cet essai clinique pilote a utilisé une la TA pour 19 patients atteints de schizophrénie avec des hallucinations auditives verbales résistantes au traitement. Les patients ont été randomisés pour recevoir soit la TA, soit le traitement habituel pendant 7 semaines, avec des évaluations effectuées avant, et 3 mois après la thérapie. La TA a significativement réduit la sévérité des hallucinations auditives (total PSYRATS : d de Cohen = 1.0 ; détresse : d = 1.2) et amélioré les symptômes dépressifs et la qualité de vie, avec des effets maintenus lors du suivi de 3 mois (du Sert et al., 2018). Le traitement habituel n'a montré aucun changement significatif dans les symptômes psychiatriques.

À la suite de ces études, l'équipe de Dr. Dumais, à Montréal, tente de comparer la TA à la TCCp. Cette étude de type essai randomisé pilote, qui compare les deux interventions pendant 9 semaines pour des patients souffrant de SRT, est en cours. Dans cet essai, les participants, qui ont continué leurs soins psychiatriques standards, ont été assignés aléatoirement soit à la TA soit à la TCCp et ont subi des évaluations cliniques avant, après, et jusqu'à 12 mois après l'intervention. Des données préliminaires sur un an rapporte que la TA et la TCCp ont toutes deux montré des améliorations statistiquement significatives à court terme dans la sévérité des hallucinations verbales auditives et des symptômes dépressifs (Dellazizzo et al., 2021). La TA a démontré des tailles d'effet plus importantes pour les hallucinations auditives verbales globales (d = 1.080), la détresse vocale (d = 0.998), et la fréquence des hallucinations auditives (d = 0.701) comparées à la TCCp (hallucinations auditives verbales : d = 0.555, détresse vocale : d = 0.434, fréquence : d = 0.339) (Dellazizzo et al., 2021). Les effets rapportés sont statistiquement significatifs. À long terme, les effets de la TA ont été maintenus jusqu'au suivi d'un an (Dellazizzo et al., 2021).

Alors que ces essais montrent des résultats prometteurs sur l'effet de la TA dans la réduction des hallucinations auditives chez les patients souffrant de schizophrénie, quelques études ont tenté d'évaluer qualitativement les verbatim (transcriptions de sessions immersives) de la TA pour mieux comprendre le processus thérapeutique. L'analyse de contenu (des sessions thérapeutiques), les entretiens semi-structurés et les questionnaires ont été couramment utilisés pour évaluer les processus thérapeutiques (Cook et al., 2017). Ces techniques sont souvent chronophages, nécessitent des ressources humaines et sont sujettes à des biais inhérents lors de l'analyse des contenus en fonction de l'approche adoptée (Hill et al., 2013). Par exemple, elles ciblent souvent un ensemble limité d'éléments, ce qui rend difficile une compréhension approfondie du processus thérapeutique intrinsèque (Szymańska et al., 2017). Des approches qualitatives telles que la phénoménologie ou la théorie ancrée sont souvent utilisées pour comprendre ce qui se passe tout au long de la session thérapeutique. Bien que les données qualitatives puissent être informatives et soient généralement vastes dans leur nature, elles manquent de leur contrepartie quantitative et rendent donc difficile la détermination de l'élément de la thérapie lié à un potentiel bon résultat (Leung, 2015). Dellazizzo et al. (2018) a mené une première analyse qualitative des séances thérapeutiques de 12 patients ayant entrepris la TA. Ils ont analysé jusqu'à 84 transcriptions de sessions immersives pour atteindre un point de saturation. Cinq thèmes émergent des réponses des patients dans un dialogue avec l'avatar dans cette analyse de contenu. Les thèmes étaient : la réponse émotionnelle aux voix, les croyances concernant les voix et la schizophrénie, la perception des patients face à leurs voix, les mécanismes d'adaptation et les aspirations. Cette étude a émis l'hypothèse que les premiers thèmes étaient liés aux cibles thérapeutiques dans la TA. Cela a conduit à une seconde étude, réalisée par Beaudoin et al. (2021), qui a évalué qualitativement 125 transcriptions de thérapie pour 18 patients (1419 minutes de thérapie) afin de mieux comprendre la dynamique entre le patient et l'avatar. Deux thèmes clés ont été identifiés pour l'avatar : les techniques de confrontation et les techniques positives. Les techniques de confrontation comprenaient un sous-ensemble de 8 sous-thèmes tandis que les techniques positives comprenaient 6 sous-thèmes. Cinq thèmes clés ont été identifiés pour le patient : la perception des patients face à leurs voix, les réponses émotionnelles, les aspirations, les mécanismes d'adaptation et croyances concernant les voix et la schizophrénie. Ces cinq thèmes comprenaient 14 sous-thèmes. L'analyse de contenu menée dans l'étude de Beaudoin a également permis de créer un ensemble de données initial pour la TA comprenant les interactions de 18 patients à travers les 28 thèmes identifiés.

Conclusion de la section

La réalité virtuelle offre une nouvelle forme d’approche thérapeutique auprès de patients en permettant une interaction ciblée avec un environnement et des avatars (Kim et al., 2020). Dans le cas des hallucinations auditives réfractaires à la médication chez les patients atteints de SRT, la TA offre une avenue prometteuse. Comme le décrit les études actuelles, il demeure toutefois une proportion de patients qui ne connaissent pas d’amélioration clinique, ou une amélioration moindre. Pour ces derniers, considérant le recensement qualitatif des processus thérapeutiques en lien avec les types d'interactions prenant place dans la TA, l'utilisation d'apprentissage machine pour analyser de façon quantitative les interactions prenant place dans la TA pourrait mener à une solution quant aux défis soulevés ci-dessus et mettre la table pour la prédiction d'une réponse thérapeutique basé davantage sur des aspects issus de l'examen mental, soit le contenu de la pensée, plutôt qu'axé uniquement sur les comportements. Cela pourrait permettre d'émettre une prédiction clinique qui pourrait potentiellement guider le thérapeute dans ses interactions avec les patients souffrant d'une SRT.

Chapitre 2 – Objectifs

Objectifs généraux

Cette thèse a pour objectif principal d'aider les thérapeutes dans l'amélioration de la réponse thérapeutique de patients souffrant de schizophrénie avec des hallucinations auditives réfractaires à la médication en intégrant les principes de l'intelligence artificielle dans la TA pour la prédiction de la réponse clinique des patients. Cinq objectifs spécifiques ont été mis sur pied afin de répondre à l'objectif principal :

(i) Utiliser l'intelligence artificielle dans la prédiction des issues cliniques des patients atteints de troubles mentaux graves et persistants ;

(ii) Évaluer les possibilités d'intégration de l'intelligence artificielle dans la psychothérapie sur les plans techniques et cliniques ;

(iii) Utiliser l'apprentissage machine dans la TA pour l'annotation automatisée des séances thérapeutiques immersives ;

(iv) Intégrer les algorithmes d'apprentissage machine pour prédire la réponse clinique des patients suivant la TA ;

(v) Connaître les facteurs prédictifs multimodaux qui pourraient bonifier la prédiction de l'issue clinique des patients.

La première section porte sur une étude transversale utilisant l'apprentissage machine afin d'identifier des facteurs prédicteurs de violence en utilisant l'apprentissage machine chez des patients présentant une complexité clinique importante. La deuxième section vise à évaluer les possibilités d'intégration de l'intelligence artificielle dans la psychothérapie sur les plans techniques et cliniques. À cette fin, deux revues de la littérature ont été réalisées. La troisième section comprend quatre études visant à utiliser l'apprentissage machine dans la thérapie par TA pour annoter automatiquement les séances thérapeutiques immersives. La quatrième section s'appuie sur les sections précédentes pour intégrer les algorithmes d'apprentissage machine en vue de prédire la réponse clinique des patients suivant la thérapie par Avatar. Finalement, la cinquième section est constituée de deux études qui offrent des perspectives intéressantes quant aux modalités additionnelles qui pourraient être explorées afin d'être intégrées dans la prédiction de la réponse clinique des patients suivant la TA.

Afin de répondre à ces objectifs, une thèse par article a été privilégiée et les objectifs de ces articles permettent de répondre aux objectifs spécifiques explorées dans le cadre de cette thèse.

Objectifs spécifiques par article

i. Utiliser l'intelligence artificielle dans la prédiction des issues cliniques des patients atteints de troubles mentaux graves et persistants

Étude 1

Hudon, A., Dellazizzo, L., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Association Between Cannabis and Violence in Community-Dwelling Patients With Severe Mental Disorders: A Cross-sectional Study Using Machine Learning. *The Journal of nervous and mental disease*, 211(2), 88–94. https://doi.org/10.1097/NMD.0000000000001604

Objectif

L'objectif de cette étude transversale a été d’identifier, grâce à une approche basée sur les données, des caractéristiques liées à la consommation de cannabis et d'autres facteurs prédictifs de comportements violents chez les patients atteints de troubles mentaux sévères. Pour ce faire, un modèle de régression avec régularisation de type *Least Absolute Shrinkage and Selection Operator* a été appliqué à une base de données comprenant 97 patients souffrant de troubles mentaux sévères et persistants, lesquels avaient rempli des questionnaires évaluant leur consommation de substances et leurs comportements violents.

ii. Évaluer les possibilités d'intégration de l'intelligence artificielle dans la

psychothérapie sur les plans techniques et cliniques

Étude 2

Hudon, A., Beaudoin, M., Phraxayavong, K., Dellazizzo, L., Potvin, S., & Dumais, A. (2021). Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms. *JMIR mental health*, 8(10), e22651. https://doi.org/10.2196/22651

Objectif

Cette étude a eu pour objectif de réaliser une revue systématique sur l'utilisation de l'apprentissage machine pour la classification automatique de textes en utilisant des bases de données de petite taille, dans les domaines de la psychiatrie, de la psychologie et des sciences sociales. L'objectif visait principalement à identifier les algorithmes disponibles et de déterminer si la classification automatique des entités textuelles est comparable à celle effectuée par des évaluateurs humains. Pour ce faire, une recherche systématique a été menée dans les bases de données électroniques de Medline, Web of Science, PsycNet (PsycINFO) et Google Scholar, depuis leurs dates de création jusqu'en 2021. Les domaines de la psychiatrie, de la psychologie et des sciences sociales ont été choisis en raison de la richesse et de la diversité des entités textuelles qu'ils comportent dans le domaine de la santé mentale. Des références supplémentaires identifiées par recoupement ont également été utilisées pour repérer d'autres études pertinentes.

Étude 3

Hudon, A., Aird, M., & La Haye-Caty, N. (2023). Deciphering the mosaic of therapeutic potential: A scoping review of neural network applications in psychotherapy enhancements. *BioMedInformatics*, 3(4), 1101–1111. https://doi.org/10.3390/biomedinformatics3040066

Objectif

L'objectif de cette étude est d'identifier les différentes applications des algorithmes de réseaux neuronaux dans le domaine de la psychothérapie. À cet effet, une revue exploratoire a été menée dans les bases de données électroniques EMBASE, MEDLINE, APA et CINAHL. La conception de cette étude s'inspire des éléments de rapport recommandés pour les revues systématiques et les méta-analyses, connus sous l'acronyme PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*). Les études retenues pour cette revue sont celles qui impliquent l'utilisation d'un algorithme de réseau neuronal dans le contexte d'une approche psychothérapeutique.

iii. Utiliser l'apprentissage machine dans la thérapie par Avatar pour l'annotation automatisée des séances thérapeutiques immersives

Étude 4

Hudon, A., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Comparing the performance of machine learning algorithms in the automatic classification of Psychotherapeutic Interactions in avatar therapy. *Machine Learning and Knowledge Extraction*, 5(3), 1119–1130. https://doi.org/10.3390/make5030057

Objectif

L'objectif de cette étude a été de comparer les performances de différents algorithmes d'apprentissage machine dans l'annotation automatique des verbatims des séances immersives pour la TA. Cinq algorithmes d'apprentissage machine issus de la bibliothèque Scikit-Learn ont été implémentés sur un ensemble de données : le classificateur à vecteurs de support, le classificateur linéaire à vecteurs de support, le Bayes naïf multinomial, les arbres décisionnel et le classificateur de perceptron multicouche. L'ensemble de données comprenait 27 types d'interactions différents survenant dans la TA, à la fois pour l'Avatar et le patient, chez 35 patients ayant participé à huit séances immersives dans le cadre de leur traitement en TA.

Étude 5

Hudon, A., Beaudoin, M., Phraxayavong, K., Dellazizzo, L., Potvin, S., & Dumais, A. (2022). Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach. *Health informatics journal*, 28(4), 14604582221142442. https://doi.org/10.1177/14604582221142442

Objectif

L'objectif principal de cette étude a été de réaliser une classification automatique des textes des interactions se produisant dans la TA. Cette étude vise également à évaluer si cette classification est comparable à celle effectuée par des codeurs humains. Pour ce faire, un classificateur linéaire de type support vectoriel a été utilisé afin d'effectuer une classification automatisée des thèmes sur les transcriptions des séances de TA, en utilisant un ensemble de données limité.

Étude 6

Hudon, A., Beaudoin, M., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Unsupervised Machine Learning Driven Analysis of Verbatims of Treatment-Resistant Schizophrenia Patients Having Followed Avatar Therapy. *Journal of personalized medicine*, 13(5), 801. https://doi.org/10.3390/jpm13050801

Objectif

L'objectif de cette étude a été de mener une analyse, en utilisant l'apprentissage machine non-supervisé, des verbatims de patients atteints de schizophrénie résistante au traitement ayant suivi la TA. Le second objectif était de comparer les clusters de données obtenus par cette analyse en apprentissage machine non supervisé avec les analyses qualitatives précédemment réalisées dans la TA. Un algorithme de k-means a été appliqué sur les verbatims des séances immersives de 18 patients souffrant de schizophrénie résistante au traitement qui ont suivi la TA, afin de regrouper les interactions entre l'avatar et le patient. Les données ont été prétraitées par vectorisation et réduction de données.

Étude 7

Hudon, A., Phraxayavong, K., Potvin, S. & Dumais A. Ensemble Methods to Optimize Automated Text Classification in Avatar Therapy. En révision dans *BioMedInformatics* (accepté avec révisions mineures, décembre 2023)

Objectif

L'objectif de cette étude a été d'évaluer l'évolution de la précision des algorithmes d'apprentissage machine pour la classification automatique des textes lorsqu'on utilise une approche d'ensemble sur les verbatims des séances immersives en TA. Un modèle d'ensemble, comprenant cinq algorithmes d'apprentissage machine, a été mis en œuvre pour effectuer la classification textuelle des interactions entre l'avatar et le patient. Les modèles inclus dans cette étude sont : le classificateur linéaire à vecteurs de support, le Bayes naïf multinomial, le classificateur de perceptron multicouche, le XGBClassifier et le modèle K-Nearest-Neighbor. La précision, le rappel, l'exactitude et le score F1 ont été comparés pour les classificateurs individuels et le modèle d'ensemble.

iv. Intégrer les algorithmes d'apprentissage machine pour prédire la réponse clinique des patients suivant la thérapie par Avatar

Étude 8

Hudon, A., Beaudoin, M., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Enhancing Predictive Power: Integrating a Linear Support Vector Classifier with Logistic Regression for Patient Outcome Prognosis in Virtual Reality Therapy for Treatment-Resistant Schizophrenia. *Journal of personalized medicine*, 13(12), 1660. https://doi.org/10.3390/jpm13121660

Objectif

L'objectif principal de cette étude a été de combiner un modèle de classification avec un modèle de régression dans le but de prédire les résultats thérapeutiques des patients, basés sur les interactions survenues lors de leur première séance immersive de thérapie en réalité virtuelle. Pour ce faire, une combinaison d'un classificateur linéaire à vecteurs de support et d'une régression logistique a été appliquée sur un ensemble de données composé de 162 verbatims issus des séances immersives de 18 patients ayant précédemment suivi une TA. Comme ensemble de données à tester, les premières séances immersives de 17 participants, inconnus de la base de données, ont été présentées au modèle combinatoire afin de prédire leur évolution clinique.

v. Connaître les facteurs prédictifs multimodaux qui pourraient bonifier la prédiction de l'issue clinique des patients

Étude 9

Hudon, A., Lammatteo, V., Rodrigues-Coutlée, S., Dellazizzo, L., Giguère, S., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Exploration of the role of emotional expression of treatment-resistant schizophrenia patients having followed virtual reality therapy: a content analysis. *BMC psychiatry*, 23(1), 420. https://doi.org/10.1186/s12888-023-04861-2

Objectif

L'objectif de cette étude a été d'identifier les émotions sous-jacentes caractérisant l'interaction entre le patient et l'Avatar lors de la TA, en analysant le contenu des transcriptions et des enregistrements audios des séances immersives. Une analyse de contenu des transcriptions et des enregistrements audios de la TA a été réalisée pour 16 patients souffrant de STR ayant suivi la TA entre 2017 et 2022, incluant 128 transcriptions et 128 enregistrements audio. Une technique de catégorisation itérative a été employée pour identifier les différentes émotions exprimées par le patient et l'Avatar au cours des séances immersives.

Étude 10

Hudon, A., Couture, J., Dellazizzo, L., Beaudoin, M., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Dyadic Interactions of Treatment-Resistant Schizophrenia Patients Having Followed Virtual Reality Therapy: A Content Analysis. *Journal of clinical medicine*, *12*(6), 2299. https://doi.org/10.3390/jcm12062299

Objectif

Cette étude a eu comme objectif d'identifier les interactions dyadiques les plus fréquentes entre le patient et l'Avatar dans la TA chez les patients souffrant de SRT. Pour atteindre cet objectif, une analyse de contenu a été menée sur 256 verbatims issus de 32 patients ayant complété la TA entre 2017 et 2022 à l'Institut universitaire en santé mentale de Montréal. Cette analyse visait à identifier les interactions dyadiques entre les patients et leur Avatar.

Chapitre 3 – RÉSULTATS

Article 1. Association Between Cannabis and Violence in Community-Dwelling Patients With Severe Mental Disorders: A Cross-sectional Study Using Machine Learning

Alexandre Hudon

Laura Dellazizzo

Kingsada Phraxayavong

Stéphane Potvin

Alexandre Dumais

Publié dans *The Journal of Nervous and Mental Disease (2023).*
Article révisé par les pairs.

Hudon, A., Dellazizzo, L., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Association Between Cannabis and Violence in Community-Dwelling Patients With Severe Mental Disorders: A Cross-sectional Study Using Machine Learning. The Journal of nervous and mental disease, 211(2), 88–94. https://doi.org/10.1097/NMD.0000000000001604

Abstract

The objective of this cross-sectional study was to identify cannabis-related features and other characteristics predictive of violence using a data-driven approach in patients with severe mental disorders (SMD). A Least Absolute Shrinkage and Selection Operator regularization regression model was used on the database consisting of 97 patients with SMD who completed questionnaires measuring substance use and violence. Cannabis use, particularly related to patients' decision to consume or time spent using, was a key predictor associated to violence. Other patterns of substance use, and personality traits were identified as strong predictors. Stable addictive patterns of cannabis use, and interpersonal issues related to cannabis/stimulant abuse were inversely correlated to violence. This study identified the effect of several predictors correlated to violence in patients suffering from SMD using a regularization regression model. Findings open the door to better identify the profiles of patients that may be more susceptible to perpetrate violent behaviours.

Keywords

machine learning, cannabis, violence, predictive modelling, severe mental disorder

Introduction

Violence is a complicated and multifaceted issue that has profound health and social effects.There are nearly 1.6 million deaths in the world being due to violence, and 10-40 times more physical injuries requiring medical attention (World Health Organization, 2014). Markedly, literature has suggested that individuals with severe mental disorders (SMD), such as schizophrenia, are associated with an increased risk of violent and non-violent crime as compared to the general population (Swanson et al., 2002). It has also been shown that those with SMD are at a greater risk of having multiple incarcerations compared to those without such disorders (Fazel, Gulati, L. Linsell, Geddes, & Grann, 2009; Fazel, Lichtenstein, Grann, Goodwin, & Långström, 2010; Fazel & Seewald, 2012). Given the multitude of deleterious effects from violence, several studies have focused on evaluating the correlates and predictors of violent behavior.

Numerous variables associated with the occurrence of violence have been identified, such as difficulties in emotion regulation (i.e., anger/hostility) and self-regulatory symptoms (impulsivity) (Menahem I. Krakowski & Czobor, 2018; M. I. Krakowski et al., 2016; Rund, 2018), which are also observed across numerous at-risk populations, such as those with psychiatric disorders (Menahem I. Krakowski & Czobor, 2018; M. I. Krakowski et al., 2016; Rund, 2018). Yet, one of the most consistent risk factors for the perpetration of violence is substance use, more so problematic use, which has also been linked to exacerbate psychiatric symptoms (e.g., psychotic symptoms, impulsivity) (Schoeler et al., 2016). While literature varies greatly, the general association between substance use and violence has long been established (Boles & Miotto, 2003; Chermack et al., 2010; Douglas, 2015; Friedman, 1999), mainly for alcohol and stimulants. Although some authors view psychiatric diagnoses as a contributing factor to violence, some literature has evidenced that violence in this specific population of SMD may result predominantly from comorbid substance use and substance use disorder (Seena Fazel, Gautam Gulati, Louise Linsell, John R. Geddes, & Martin Grann, 2009). More particularly, cannabis use is amid the substances being far less investigated in association with violence, notably in patients with SMD (Dellazizzo et al., 2019), with most studies being conducted in the general population (Dellazizzo et al., 2020). This is not surprising as public perceptions have tended to disregard the potential effects of cannabis, mainly on harms towards others.

Nevertheless, in terms of clinical implications, cannabis accounts for a substantial proportion of disease burden (Degenhardt, Ferrari, & Hall, 2017; Imtiaz et al., 2016) and for around half of all first-time admission to specialist drug treatment worldwide (United Nations Office on Drugs and Crime (UNODC), 2018). Cannabis is one of the most used addictive substances in SMD samples, with rates surpassing those of the general populations, and prior studies have reported high rates of cannabis use disorder as well (Cantor-Graae, Nordström, & McNeil, 2001; Green, Young, & Kavanagh, 2005; Mueser et al., 1990). This may reflect an attempt to cope with psychological distress (e.g., negative affective symptoms) or relieve the side effects of medication (e.g., antipsychotics) through cannabis use (e.g., self-medication) (Goswami, Mattoo, Basu, & Singh, 2004). Recently, there has been a change surrounding societal and legal viewpoints on cannabis suggesting shifting public attitudes towards the perceived safety and social acceptability of its use (Leung, Chiu, Stjepanović, & Hall, 2018). Some evidence has also suggested changes in prevalence and frequency of use in some samples (Hasin et al., 2016; Lake et al., 2019; Melchior et al., 2019; Salas-Wright et al., 2017). Moreover, cannabis potency levels, as measured by Δ9-tetrahydrocannabinol (THC) content in relation to cannabidiol (CBD) content, has substantially increased 3- to 4-fold (Chandra et al., 2019; Mahamad, Wadsworth, Rynard, Goodman, & Hammond, 2020). The majority of the psychoactive and emotion-related effects as well as addictive properties of cannabis are due to THC (Atakan, 2012). The latter exerts a range of temporary and dose-dependent effects by acting on the central nervous system primarily via cannabinoid receptor type 1 (CB1), which mediates inhibitory action on the release of difference neurotransmitters (Lambert & Fowler, 2005; Pertwee, 2006, 2008a, 2008b). Whereas available data advocates more harms associated with cannabis use in individuals with psychiatric disorders than benefits, the deleterious effects of cannabis use remain a matter of debate, especially pertaining to violence. Although the association between cannabis and violence remains heterogeneous (i.e. (Dharmawardene & Menkes, 2017; Fergusson & Horwood, 1997; Haggard-Grann, Hallqvist, Langstrom, & Moller, 2006; Macdonald et al., 2003; Norström & Rossow, 2014; Wei, Loeber, & White, 2004)), literature has identified several possible mechanisms by which cannabis use may be associated with violence, including symptoms related to intoxication and withdrawal, poor clinical outcomes, worsening of psychotic symptoms, interaction with developmental predispositions for aggression, etc (Moore & Stuart, 2005).

Concordantly with these mechanisms, there is early meta-analytical evidence pointing towards a positive association cannabis use/misuse and the perpetration of violence in adult individuals with SMD (Dellazizzo et al., 2019). The meta-analysis included 12 final articles amounting to a total of 3873 subjects with results showing a moderate association between cannabis use and violence in individuals with SMD (OR=3.02, CI=2.01-4.54). Interestingly, the association was significantly higher for cannabis misuse in comparison to cannabis use (OR=5.8, CI=3.27-10.28 versus OR=2.04, CI=1.36-3.05). Nevertheless, results were based solely on a few studies with a variety of methodological issues as little research has aimed to further investigate the relationship between cannabis and violence in SMD samples in contrast to the general population. Hence, most studies were of retrospective or cross-sectional nature and did not account for important confounding factors (e.g., sociodemographic and clinical data, other substance use). Additionally, the database was characterized by high heterogeneity, which may have been partly due to the studies displaying a variety of definitions for violence as well as cannabis use and assessment methods. Therefore, it is difficult to establish which precise key elements related to cannabis use, such as frequency of use, THC content, use patterns, increase the risk of violence. Hence, studies on the subject have rarely controlled for the impact of confounding factors when assessing for correlations between cannabis use and violence. In this sense and following the liberalization for cannabis policies, further studies are required to enlighten the nature of the association between cannabis and the perpetration of violence in patients with SMD, which we aimed to do by employing machine learning (Boden & Spittlehouse, 2019).

The primary objective of our cross-sectional study was to identify cannabis-related features (i.e., characteristics) of patients with SMD that are predictive of violence using a data-driven approach. The secondary objective of this study was to identify other features associated to violence and their specific impact on violence as aggravating factors. It is hypothesized that consuming cannabis in the past 12 months is linked to an increase of violence. Other features associated such as anti-social personality traits and stimulant use could also contribute to violent behaviors.

Methods

Participants

A cross-sectional study comprising 104 subjects recruited for the Cannabis Violence project (CANVI Project) between June 2018 and January 2020 was performed. The CANVI Project conducted at the Research Center of the (blinded for peer revision) attempts to link evidence related to cannabis use and violent behaviors in patients suffering from SMD. Subjects were included in the study if they consumed cannabis at least once in their lifetime, were 18 years of age or older, were diagnosed with a severe psychiatric disorder and were not homeless (to ensure possibility of follow-up). They were recruited in three ways, that is either via self-referrals, referrals by their treating clinical team or via the Signature Bank of the (blinded for peer revision).

Measures

Upon approval to join the project, participants completed several self-reported questionnaires assessing socio-demographic characteristics, impulsivity, psychiatric disorders, psychotic symptoms, cannabis use as well as other substance use and violent behaviors. Table 1 shows a description of questionnaires as well as validity and reliability indices from the literature from studies with comparable samples. These questionnaires were selected based on their use in the field on topics relevant to psychosis, violence and cannabis.

Cannabis use

The Cannabis Use Problems Identification Test (CUPIT) is a self-reported questionnaire used to assess the frequency, quantity, and patterns of participants' cannabis consumption (Bashford, Flett & Copeland, 2010).

Violence

The MacArthur Community Violence Instrument (MacCVI) was used to assess participants' prior violent behaviors and violent acts committed against them (Monahan, Steadman, Silver, Appelbaum, Robbins, Mulvey et al., 2001). MacCVI items were rated for each participant on the Cormier-Lang Criminal History Score for Violent Offences to obtain a numerical score associated to violent offences (Quinsey et al., 1998). This was performed to account for the frequency and severity of violent offences.

-Please Insert Table 1 Here –

Data pre-processing and cross-validation

Data was pre-processed prior to entering the model. Participants who completed less than 25% of the questionnaires in addition to participants with missing MacCVI scores were removed from the dataset. Missing data was accounted for by using a mean imputation technique in which the mean of the observed values for each variable was computed and missing values for that same variable were imputed by this mean (Austin, White, Lee & van Buuren, 2021). Cross-validation of the model was conducted using a 10-fold algorithm (Anguita, Ghelardoni, Ghio, Oneto & Ridella, 2012).

Variables

The dependent variable used in the model was the numerical score of the Cormier-Lang Criminal History Score for Violent Offences. All the questionnaires' individual items (and not the total scores of each questionnaire) were the independent variables (also known as features) used in the model. A dataset containing the independent variables and dependent variable for each subject was created and presented to the model. Multicollinearity was addressed with the regression model.

Regression model

A regression model was programmed in Python 3.9. To account for overfitting and to include all independent variables in the analysis, a Least Absolute Shrinkage and Selection Operator (LASSO) regularization regression model was used from the Sci-kit learn library (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel et al., 2011). This type of algorithm, which is a type of linear regression, is better suited for datasets in which there is a high chance of multicollinearity across the independent variables (Ranstam & Cook, 2018). It uses a method, termed shrinkage, where data values are shrunk towards a central point; the coefficient of variables that are not strong enough to be included in the prediction is therefore reduced to 0 (Tibshirani, 2011). In this study, independent variables with high coefficients show a positive influence on violence whereas negative coefficients show variables that may be considered as being protective factors. The algorithm's performance was assessed with the R2 test score and the variance trade-off of the model was assessed by calculating the mean squared train and test errors. A R2 test score of 0% defines a model that does not explain any of the variations in the dependant variable around its mean whereas a R2 test score of 100% represents a model that explains all the variation in the response variable around its mean (Chicco, Warrens & Jurman, 2021). The hyperparameters for the model were identified using the GridSearchCV algorithm from the Sci-kit learn library (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel et al., 2011).

Results

Out of 104 participants, data for 97 participants were included in this study since 7 were excluded in the pre-processing phase. Characteristics of patients are presented in Table 2. The algorithm achieved a mean performance R2 score of 0.41 (min=0.37, max=0.42) with a mean squared train and test scores of X and Y respectively. The hyperparameters for the model are alpha=0.316 and the default parameters. The retained predictive features were statistically significant (p-value <0.01).

-Please Insert Table 2 Here –

Cannabis-related features and their correlations with violent behaviors are presented in Table 3. As it may be observed, using cannabis in the past 12 months after deciding that it would be better to withhold use, having difficulties to go through a day without using and considering the time spent using cannabis to be problematic are strong positive predictors related to cannabis for violent behaviors.

Other features having been identified as having a positive correlation with violence are also displayed in Table 3. Items related to substance abuse, mainly for cannabis, stimulants and cocaine, have been found to be strong predictive features. Notably, stimulant use in the past 12 months associated to withdrawal symptoms was identified as the strongest predictor. Past violent behaviors and personality traits, such as conducting illegal acts, antecedents of temper tantrums after the age of 16, stealing and voluntarily destroying belongings were likewise strong predictors following substance abuse-related features.

-Please Insert Table 3 Here -

Features showing a negative predictive correlation with violence may be found in Table 4. Chronic substance use patterns in participants with poor judgement (e.g., using in the presence of an imminent danger) were inversely associated with violent behaviors. Coefficients found for these items appear to be negligible in the prediction of violent behaviors in comparison to the positive predictors found since they were in most part of smaller magnitude. Negative predictors included not having conducted illicit behaviors, such as lying before the age of 15, having never done anything illegal, and having never experienced disciplinary issues at school or with the law. Prior lifetime cannabis and alcohol use despite imminent danger were identified as the strongest negative predictors, closely followed by chronic cannabis use in the past 12 months.

-Please Insert Table 4 Here -

Discussion

The goal of this study was primarily to identify cannabis related characteristics in patients with SMD that were predictive of violent behaviors and, secondly, to identify further predictors of violence. The LASSO algorithm's predictive performance of R2=0.41 is consistent to predictive scores found in the literature for predicting different human behaviors (Ferguson, 2009 ; Hamilton, Ghert & Simpson, 2015).

First, cannabis use was found to be a key predictor associated to violent behaviors, which is not surprising considering the number of studies showing a relationship between cannabis and violence (Dellazizzo, Potvin, Beaudoin, Luigi, Dou, Giguère & Dumais, 2019). Using cannabis may increase aggressiveness, paranoia and angriness (Courts, Maskill, Gray & Glue, 2016). Cannabis is also known to be a risk factor for negative behavioral outcomes (such as the negative symptoms seen in schizophrenia) for patients with psychiatric disorders (Shah, Chand, Bandawar, Benegal & Murphy, 2017). Our results suggest that cannabis use related to patients' decision to consume or the time spent using the substance was found to be positively correlated to violent gestures. Markedly, patterns of cannabis use in the last 12 months, including using cannabis after having decided to withhold use or finding it difficult to go through a day without using were also identified as strong predictors. This follows a study published in 2020 having examined the behaviours of adolescents and young adults regarding cannabis use. Authors identified that being deprived of using cannabis while being dependent elicited irritability and yielded to more violent gestures (Miller, Ipeku & Oberbarnscheidt, 2020). Moreover, a focused review noted a moderate association between cannabis use and violence in the 4 meta-analyses retrieved with results on samples including youths, intimate partners and individuals with SMD (Dellazizzo, Potvin, Athanassiou & Dumais, 2020)). Such results were confirmed in our study considering that cannabis use and violent traits associated to SMD were also identified including items linked to antisocial personality traits. Accordingly, several studies have shown that conduct disorder traits prior to the age of 15 years and antisocial personality disorder traits prior to the age of 18 years was associated to an increase in violent behaviors (Ilomäki, Viilo, Hakko, Marttunen, Mäkikyrö & Räsänen, 2006 ; DeLisi, Drury & Elbert, 2019). Considering that substance abusers, especially related to cannabis abuse, has been reported to be correlated with an increase in violent behaviors as compared to non-users and, on the other hand, that violent traits at a young age are also related to violent behaviors this may explain why top predictors identified in our study fall within these categories (Pickard & Fazel, 2013). Several studies on SMD populations have indicated that long-term cannabis use can yield to cognitive variations characterized by a lack of inhibition (Lovell, Bruno, Johnston, Matthews, McGregor, Allsop & Lintzeris, 2018; Jenkins & Kohkhar, 2021). For instance, as seen in studies with elderly patients who use cannabis (displaying a severe decrease in inhibition), it may be hypothesized that patients with SMD with prefrontal cortical atrophy (e.g., as seen in patients suffering from schizophrenia) combined with cannabis use may be associated to increased violent behaviors (Albaugh et al., 2021).

Second, our study identified certain cannabis predictors as protective factors for violence. Chronic and stable addictive patterns of cannabis use, such as using the substance in the morning or using daily were inversely correlated to violent behaviours. Chronic and stable cannabis use (ex.: daily cannabis use in the past 3 months) might be related to fewer withdrawal symptoms and lack of motivation (Pacheco-Colón, Limia & Gonzalez, 2018), which appears to decrease violent behaviors.

Third, other patterns of substance use and personality traits (notably antisocial personality traits) were identified as strong predictors of violence. Irritability and frustration leading to stimulant withdrawal identified as a strong predictor in our model in accordance with literature on the subject (Gilchrist, Dennis, Radcliffe, Henderson, Howard & Gadd, 2019). Similarly, cluster B personality disorders and cocaine use have been often correlated to violent behaviour, which is consistent with our findings (Esbec & Echeburúa, 2010). In terms of protective factors, interpersonal issues associated to cannabis or stimulant abuse negatively predicted violent acts in comparison to cocaine. Notably, alcohol and cocaine use have often been found to be comorbid with personality disorders, which may explain the association with the interpersonal issues (Parmar & Kaloiya, 2018). Outside of the realm of substance abuse, the absence of problems with the law, lack of disciplinary issues at school or an absence of family conflicts were naturally negatively correlated with violence in accordance with prior literature (Borowsky, Ireland & Resnick, 2002). Nevertheless, some results related to the negative predictors of violence went against the findings of many studies. These factors included cocaine use with withdrawal symptoms and stimulants use with interpersonal problems. This could be explained by the attempt of solving lifetime cocaine and stimulants use addictions, which might slightly reduce the chance of more recent violent behaviours. However, these findings may be negligible as the correlation coefficients of negative predictors in comparison to positive predictors are leading to a very low yield in the overall prediction. Consequently, the model places a greater emphasis on the predictors with a positive coefficient to predict the outcome.

Limitations

This study has few limitations that should be noted. Since this study was cross-sectional in nature, the potential analysis of the temporal association between cannabis use and violence cannot be investigated. Moreover, self-reported violent behaviors can result in incorrect MacCVI scores, which could undermine the strength of the relationship between the identified predictors and violent behaviors. However, prior studies have shown self-reported violence to be sufficiently reliable as study outcomes (Monahan, Steadman, Silver, Appelbaum, Robbins, Mulvey et al., 2001). Lastly, the small number of participants as compared to the number of tested variables undermines the possibility to achieve a greater predictive R2 score despite having reached an acceptable predictive score. Future longitudinal studies on the subject are needed to better understand the role of cannabis use in the prediction of violent behaviours.

Conclusions

This study identified several predictors correlated to violent behaviors in patients suffering from SMD and comorbid cannabis use. Specific cannabis abuse traits, especially those involving practical judgement, tended to be associated with violent acts as compared to patterns of chronic use. Other patterns of substance abuse of certain drugs, such as stimulants, cocaine and, to a lesser extent, alcohol, were identified as potential predictors. This study opens the door to better identify the profiles of patients that may be more susceptible to perpetrate violent behaviours. Also, knowledge on the positive predictors may help prevent future violent acts. Clinically, this study offers an opportunity to distinguish key predictors among patterns that could help a clinician to better assess specific aspects of their patients when investigating for potential violent behaviours. This could be attained by better profiling patterns of cannabis use in patients suffering from SMD. Further studies are needed on the subject to include a wider array of patient profiles.

Disclosure

Funding sources: L. D. is a holder of scholarships from the Fonds de Recherche du Québec en Santé. S. P. is a holder of the Eli Lilly Canada Chair on schizophrenia research. A. D. is holder of a Junior 2 salary award from the Fonds de Recherche du Québec en Santé. A. H. is holder of a bursary from the Fonds d' excellence en recherhe Apogée Canada. These fund sources had no

role in the study design, collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication.

Contributions: A. H., L. D., and K. P. designed the study. A. H., L. D., S. P., and A. D. conducted literature searches and provided summaries of previous research studies. A. H. conducted the statistical analysis. A. H., L. D., and A. D. wrote the first draft of the manuscript and all authors contributed to and have approved the final manuscript.

Ethics institutional review board statement: This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients.

References

Albaugh, M., Ottino-Gonzalez, J., Sidwell, A., Lepage, C., Juliano, A., & Owens, M., et al. (2021). Association of Cannabis Use During Adolescence With Neurodevelopment. JAMA Psychiatry, 78(9), 1031. https://doi.org/10.1001/jamapsychiatry.2021.1258

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). The 'K'in K-fold cross validation. In 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) (pp. 441-446). i6doc. com publ.

Atakan, Z. (2012). Cannabis, a complex plant: different compounds and different effects on individuals. Ther Adv Psychopharmacol, 2(6), 241-254. doi:10.1177/2045125312457586

Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing Data in Clinical Research: A Tutorial on Multiple Imputation. The Canadian journal of cardiology, 37(9), 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

Bashford, J., Flett, R., & Copeland, J. (2010). The Cannabis Use Problems Identification Test (CUPIT): development, reliability, concurrent and predictive validity among adolescents and adults. Addiction (Abingdon, England), 105(4), 615–625. https://doi.org/10.1111/j.1360-0443.2009.02859.x

Boden, J. M., & Spittlehouse, J. K. (2019). What we know, and don't know, about cannabis, psychosis and violence. The New Zealand medical journal, 132(1499), 76–77.

Boles, S. M., & Miotto, K. (2003). Substance abuse and violence: A review of the literature. Aggression and violent behavior, 8(2), 155-174.

Borowsky, I. W., Ireland, M., & Resnick, M. D. (2002). Violence risk and protective factors among youth held back in school. Ambulatory pediatrics: the official journal of the Ambulatory Pediatric Association, 2(6), 475–484. https://doi.org/10.1367/15394409(2002)002<0475:vrapfa>2.0.co;2

Cantor-Graae, E., Nordström, L. G., & McNeil, T. F. (2001). Substance abuse in schizophrenia: a review of the literature and a study of correlates in Sweden. Schizophr Res, 48(1), 69-82. doi:10.1016/s0920-9964(00)00114-6

Chandra, S., Radwan, M. M., Majumdar, C. G., Church, J. C., Freeman, T. P., & ElSohly, M. A. (2019). New trends in cannabis potency in USA and Europe during the last decade (2008-2017). Eur Arch Psychiatry Clin Neurosci, 269(1), 5-15. doi:10.1007/s00406-019-00983-5

Chermack, S. T., Grogan-Kaylor, A., Perron, B. E., Murray, R. L., De Chavez, P., & Walton, M. A. (2010). Violence among men and women in substance use disorder treatment: A multi-level event-based analysis. Drug and Alcohol Dependence, 112(3), 194-200.

Chicco, D., Warrens, M., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj Computer Science, 7, e623. https://doi.org/10.7717/peerj-cs.623

Courts, J., Maskill, V., Gray, A., & Glue, P. (2016). Signs and symptoms associated with synthetic cannabinoid toxicity: systematic review. Australasian psychiatry : bulletin of Royal Australian and New Zealand College of Psychiatrists, 24(6), 598–601. https://doi.org/10.1177/1039856216663733

de Meneses-Gaya, C., Zuardi, A., Loureiro, S., & Crippa, J. (2009). Alcohol Use Disorders Identification Test (AUDIT): An updated systematic review of psychometric properties. Psychology & Neuroscience, 2(1), 83-97. https://doi.org/10.3922/j.psns.2009.1.12

Degenhardt, L., Ferrari, A. J., & Hall, W. D. (2017). Chapter 10 - The Global Epidemiology and Disease Burden of Cannabis Use and Dependence. In V. R. Preedy (Ed.), Handbook of Cannabis and Related Pathologies (pp. 89-100). San Diego: Academic Press.

DeLisi, M., Drury, A. J., & Elbert, M. J. (2019). The etiology of antisocial personality disorder: The differential roles of adverse childhood experiences and childhood psychopathology. Comprehensive psychiatry, 92, 1–6. https://doi.org/10.1016/j.comppsych.2019.04.001

Dellazizzo, L., Potvin, S., Giguère, C., Berwald, M., Dugré, J., & Dumais, A. (2017). The psychometric properties of the Life History of Aggression evaluated in patients from a psychiatric emergency setting. Psychiatry Research, 257, 485-489. https://doi.org/10.1016/j.psychres.2017.08.031

Dellazizzo, L., Potvin, S., Beaudoin, M., Luigi, M., Dou, B. Y., Giguère, C., & Dumais, A. (2019). Cannabis use and violence in patients with severe mental illnesses: A meta-analytical investigation. Psychiatry Res, 274, 42-48. doi:10.1016/j.psychres.2019.02.010

Dellazizzo, L., Potvin, S., Dou, B. Y., Beaudoin, M., Luigi, M., Giguère, C.-É., & Dumais, A. (2020). Association Between the Use of Cannabis and Physical Violence in Youths: A Meta-

Analytical Investigation. American Journal of Psychiatry, appi.ajp.2020.19101008. doi:10.1176/appi.ajp.2020.19101008

Dellazizzo, L., Potvin, S., Athanassiou, M., & Dumais, A. (2020). Violence and Cannabis Use: A Focused Review of a Forgotten Aspect in the Era of Liberalizing Cannabis. Frontiers in psychiatry, 11, 567887. https://doi.org/10.3389/fpsyt.2020.567887

Dharmawardene, V., & Menkes, D. B. (2017). Violence and self-harm in severe mental illness: inpatient study of associations with ethnicity, cannabis and alcohol. Australas Psychiatry, 25(1), 28-31. doi:10.1177/1039856216671650

Douglas, K. S. (2015). Addiction and Violence Risk. The Encyclopedia of Clinical Psychology.

Esbec, E., & Echeburúa, E. (2010). Violence and personality disorders: clinical and forensic implications. Actas espanolas de psiquiatria, 38(5), 249–261.

Fazel, S., Gulati, G., Linsell, L., Geddes, J. R., & Grann, M. (2009). Schizophrenia and violence: systematic review and meta-analysis. PLoS Med, 6(8), e1000120. doi:10.1371/journal.pmed.1000120

Fazel, S., Lichtenstein, P., Grann, M., Goodwin, G. M., & Långström, N. (2010). Bipolar Disorder and Violent Crime: New Evidence From Population-Based Longitudinal Studies and Systematic Review. Archives of general psychiatry, 67(9), 931-938. doi:10.1001/archgenpsychiatry.2010.97 %J Archives of General Psychiatry

Fazel, S., & Seewald, K. (2012). Severe mental illness in 33,588 prisoners worldwide: systematic review and meta-regression analysis. Br J Psychiatry, 200(5), 364-373. doi:10.1192/bjp.bp.111.096370

Fergusson, D. M., & Horwood, L. (1997). Early onset cannabis use and psychosocial adjustment in young adults. Addiction, 92(3), 279-296.

Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. Professional Psychology: Research And Practice, 40(5), 532-538. https://doi.org/10.1037/a0015808

Friedman, A. S. (1999). Substance use/abuse as a predictor to illegal and violent behavior: A review of the relevant literature. Aggression and violent behavior, 3(4), 339-355.

Gilchrist, G., Dennis, F., Radcliffe, P., Henderson, J., Howard, L. M., & Gadd, D. (2019). The interplay between substance use and intimate partner violence perpetration: A meta-ethnography. The International journal on drug policy, 65, 8–23. https://doi.org/10.1016/j.drugpo.2018.12.009

Goswami, S., Mattoo, S. K., Basu, D., & Singh, G. (2004). Substance-abusing schizophrenics: do they self-medicate? Am J Addict, 13(2), 139-150. doi:10.1080/10550490490435795

Green, B., Young, R., & Kavanagh, D. (2005). Cannabis use and misuse prevalence among people with psychosis. The British Journal of Psychiatry, 187(4), 306-313.

Haggard-Grann, U., Hallqvist, J., Langstrom, N., & Moller, J. (2006). The role of alcohol and drugs in triggering criminal violence: a case-crossover study*. Addiction, 101(1), 100-108. doi:10.1111/j.1360-0443.2005.01293.x

Hamilton, D. F., Ghert, M., & Simpson, A. H. (2015). Interpreting regression models in clinical outcome studies. Bone & joint research, 4(9), 152–153. https://doi.org/10.1302/2046-3758.49.2000571

Hasin, D. S., Kerridge, B. T., Saha, T. D., Huang, B., Pickering, R., Smith, S. M., . . . Grant, B. F. (2016).

Prevalence and Correlates of DSM-5 Cannabis Use Disorder, 2012-2013: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions-III. Am J Psychiatry, 173(6), 588-599. doi:10.1176/appi.ajp.2015.15070907

Ilomäki, E., Viilo, K., Hakko, H., Marttunen, M., Mäkikyrö, T., & Räsänen, P. (2006). Familial risks, conduct disorder and violence. European child & adolescent psychiatry, 15(1), 46-51.

Imtiaz, S., Shield, K. D., Roerecke, M., Cheng, J., Popova, S., Kurdyak, P., . . . Rehm, J. (2016). The burden of disease attributable to cannabis use in Canada in 2012. Addiction, 111(4), 653-662. doi:10.1111/add.13237

Jenkins, B., & Khokhar, J. (2021). Cannabis Use and Mental Illness: Understanding Circuit Dysfunction Through Preclinical Models. Frontiers In Psychiatry, 12. https://doi.org/10.3389/fpsyt.2021.597725

Krakowski, M. I., & Czobor, P. (2018). Distinctive profiles of traits predisposing to violence in schizophrenia and in the general population. Schizophrenia research, 202, 267-273. doi:https://doi.org/10.1016/j.schres.2018.07.008

Krakowski, M. I., De Sanctis, P., Foxe, J. J., Hoptman, M. J., Nolan, K., Kamiel, S., & Czobor, P. (2016). Disturbances in Response Inhibition and Emotional Processing as Potential Pathways to Violence in Schizophrenia: A High-Density Event-Related Potential Study. Schizophr Bull, 42(4), 963-974. doi:10.1093/schbul/sbw005

Lake, S., Kerr, T., Werb, D., Haines-Saah, R., Fischer, B., Thomas, G., . . . Milloy, M.-J. (2019). Guidelines for public health and safety metrics to evaluate the potential harms and benefits of cannabis regulation in Canada. Drug and Alcohol Review, 38(6), 606-621. doi:10.1111/dar.12971

Lambert, D. M., & Fowler, C. J. (2005). The endocannabinoid system: drug targets, lead compounds, and potential therapeutic applications. J Med Chem, 48(16), 5059-5087. doi:10.1021/jm058183t

Leung, J., Chiu, C. Y. V., Stjepanović, D., & Hall, W. (2018). Has the Legalisation of Medical and Recreational Cannabis Use in the USA Affected the Prevalence of Cannabis Use and Cannabis Use Disorders? Curr Addict Rep, 5(4), 403-417. doi:10.1007/s40429-018-0224-9

Lovell, M. E., Bruno, R., Johnston, J., Matthews, A., McGregor, I., Allsop, D. J., & Lintzeris, N. (2018). Cognitive, physical, and mental health outcomes between long-term cannabis and tobacco users. Addictive behaviors, 79, 178–188. https://doi.org/10.1016/j.addbeh.2017.12.009

Macdonald, S., Anglin-Bodrug, K., Mann, R. E., Erickson, P., Hathaway, A., Chipman, M., & Rylett, M. (2003). Injury risk associated with cannabis and cocaine use. Drug and Alcohol Dependence, 72(2), 99-115. doi:http://dx.doi.org/10.1016/S0376-8716(03)00202-3

Mahamad, S., Wadsworth, E., Rynard, V., Goodman, S., & Hammond, D. (2020). Availability, retail price and potency of legal and illegal cannabis in Canada after recreational cannabis legalization. 39(4), 337-346. doi:10.1111/dar.13069

Melchior, M., Nakamura, A., Bolze, C., Hausfater, F., El Khoury, F., Mary-Krause, M., & Azevedo Da Silva, M. (2019). Does liberalisation of cannabis policy influence levels of use in adolescents and young adults? A systematic review and meta-analysis. BMJ Open, 9(7), e025880. doi:10.1136/bmjopen-2018-025880

Miller, N. S., Ipeku, R., & Oberbarnscheidt, T. (2020). A Review of Cases of Marijuana and Violence. International journal of environmental research and public health, 17(5), 1578*.

https://doi.org/10.3390/ijerph17051578

Monahan J, Steadman HJ, Silver E, Appelbaum PS, Clark Robbins P, Mulvey EP et al (2001) Rethinking risk assessment. The MacArthur study of mental disorder and violence. Oxford University Press, New York

Moore, T. M., & Stuart, G. L. (2005). A review of the literature on marijuana and interpersonal violence. Aggression and violent behavior, 10(2), 171-192.

Mueser, K. T., Yarnold, P. R., Levinson, D. F., Singh, H., Bellack, A. S., Kee, K., . . . Yadalam, K. G. (1990). Prevalence of substance abuse in schizophrenia: demographic and clinical correlates. Schizophr Bull, 16(1), 31-56. doi:10.1093/schbul/16.1.31

Norström, T., & Rossow, I. (2014). Cannabis use and violence: Is there a link? Scandinavian journal of public health, 42(4), 358-363.

Osório, F., Loureiro, S., Hallak, J., Machado-de-Sousa, J., Ushirohira, J., & Baes, C. et al. (2019). Clinical validity and intrarater and test–retest reliability of the Structured Clinical Interview for DSM-5 – Clinician Version (SCID-5-CV). Psychiatry And Clinical Neurosciences, 73(12), 754-760. https://doi.org/10.1111/pcn.12931

Pacheco-Colón, I., Limia, J. M., & Gonzalez, R. (2018). Nonacute effects of cannabis use on motivation and reward sensitivity in humans: A systematic review. Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors, 32(5), 497–507*. https://doi.org/10.1037/adb0000380

Parmar, A., & Kaloiya, G. (2018). Comorbidity of Personality Disorder among Substance Use Disorder Patients: A Narrative Review. Indian journal of psychological medicine, 40(6), 517–527*. https://doi.org/10.4103/IJPSYM.IJPSYM_164_18

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825

Peralta, V., & Cuesta, M. J. (1994). Psychometric properties of the positive and negative syndrome scale (PANSS) in schizophrenia. Psychiatry research, 53(1), 31–40*. https://doi.org/10.1016/0165-1781(94)90093-0

Pertwee, R. G. (2006). Cannabinoid pharmacology: the first 66 years. Br J Pharmacol, 147 Suppl 1(Suppl 1), S163-171. doi:10.1038/sj.bjp.0706406

Pertwee, R. G. (2008a). The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: delta9-tetrahydrocannabinol, cannabidiol and delta9-tetrahydrocannabivarin. Br J Pharmacol, 153(2), 199-215. doi:10.1038/sj.bjp.0707442

Pertwee, R. G. (2008b). Ligands that target cannabinoid receptors in the brain: from THC to anandamide and beyond. Addict Biol, 13(2), 147-159. doi:10.1111/j.1369-1600.2008.00108.x

Pickard, H., & Fazel, S. (2013). Substance abuse as a risk factor for violence in mental illness: some implications for forensic psychiatric practice and clinical ethics. Current opinion in psychiatry, 26(4), 349–354. https://doi.org/10.1097/YCO.0b013e328361e798

Quinsey VL, Harris GT, Rice ME, Cormier CA (1998). Violent Offenders: Appraising and Managing Risk. Washington, DC: American Psychological Association

Ranstam, J., & Cook, J. A. (2018). LASSO regression. Journal of British Surgery, 105(10), 1348-1348.

Rund, B. R. (2018). A review of factors associated with severe violence in schizophrenia. Nordic journal of psychiatry, 72(8), 561-571.

Salas-Wright, C. P., Vaughn, M. G., Cummings-Vaughn, L. A., Holzer, K. J., Nelson, E. J., AbiNader, M., & Oh, S. (2017). Trends and correlates of marijuana use among late middle-aged and older adults in the United States, 2002-2014. Drug and Alcohol Dependence, 171, 97-106. doi:10.1016/j.drugalcdep.2016.11.031

Schoeler, T., Monk, A., Sami, M. B., Klamerus, E., Foglia, E., Brown, R., . . . Bhattacharyya, S. (2016). Continued versus discontinued cannabis use in patients with psychosis: a systematic review and meta-analysis. The Lancet Psychiatry, 3(3), 215-225.

Shah, D., Chand, P., Bandawar, M., Benegal, V., & Murthy, P. (2017). Cannabis induced psychosis and subsequent psychiatric disorders. Asian journal of psychiatry, 30, 180–184. https://doi.org/10.1016/j.ajp.2017.10.003

Shirinbayan, P., Salavati, M., Soleimani, F., Saeedi, A., Asghari-Jafarabadi, M., Hemmati-Garakani, S., & Vameghi, R. (2020). The Psychometric Properties of the Drug Abuse Screening Test. Addiction & health, 12(1), 25–33. https://doi.org/10.22122/ahj.v12i1.256

Swanson, J. W., Swartz, M. S., Essock, S. M., Osher, F. C., Wagner, H. R., Goodman, L. A., . . . Meador, K. G. (2002). The social–environmental context of violent behavior in persons treated for severe mental illness. American Journal of Public Health, 92(9), 1523-1531.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(3), 273-282.

United Nations Office on Drugs and Crime (UNODC). (2018). World Drug Report 2018 United Nations publication.

Wei, E. H., Loeber, R., & White, H. R. (2004). Teasing apart the developmental associations between alcohol and marijuana use and violence. Journal of Contemporary Criminal Justice, 20(2), 166-183.

World Health Organization. (2014). Global status report on violence prevention 2014. Geneva, Switzerland. Retrieved from www.who.int/violence_injury_prevention/violence/status_report/2014

Figures and Tables

Table 1. Description of questionnaires

Questionnaires	Number of items	Interval measured	Psychometrics
Sociodemographic (DEMO)	21	Lifetime	N/A
Drug Abuse Screening Test (DAST)	10	Last 12 months	Cronbach's alpha coefficient : 0.93 (excellent) Excellent validity (Shirinbayan, Salavati, Soleimani, Saeedi, Asghari-Jafarabadi, Hemmati-Garakani & Vameghi, 2020)
Brown-Goodwin History of Aggression (BGH)	11	Lifetime	Cronbach's alpha varied between 0.83 and 0.89 test-retest reliability coefficient: 0.80 (Dellazizzo, Potvin, Giguère,Berwald, Dugré & Dumais, 2017)
Alcohol Use Disorders Identification Test (AUDIT)	10	During the last year	High internal consistency (alpha of 0.94) Adequate test-retest reliability (de Meneses-Gaya, Zuardi, Loureiro & Crippa, 2009)
Positive And Negative Syndrome Scale (PANSS)	16	Actual	Positive and negative syndromes showed factorial validity, but they were not sufficient to account for the entirety of symptoms of schizophrenia

			(Peralta & Cuesta, 1994)
MacArthur Community Violence Instrument (MacCVI)	22	Lifetime, past 12 months	Interrater reliability for all key variables : kappa coefficients (alpha>0.80) (excellent) Kappa coefficient for violence screening ranged between 0.85 and 0.98. (Monahan, Steadman, Silver, Appelbaum, Robbins, Mulvey et al., 2001)
Cannabis Use Problems Identification Test (CUPIT)	16	Lifetime, past 12 months, past 3 months	Test–retest reliability coefficient: 0.89–0.99 (excellent) Internal consistency reliability: Cronbach's alphas ranged from 0.79 to 0.92. (excellent) (Bashford, Flett & Colepand, 2010)
Structured Clinical Interview for DSM Disorders (SCID 5: ASPD, BPD, MDD, Mania, Substance use, Psychosis)	303	Lifetime	The percentage of positive agreement between the interview and clinical diagnoses ranged between 73% and 97% The diagnostic sensitivity/specificity were >70% (Osório, Loureiro, Hallak, Machado-de-Sousa, Ushirohira & Baes, et al., 2019)

Table 2. Sociodemographic characteristics of participants

Characteristics	Average or Number of participants

Age in years, average(min-max)	40.6 (19-69)
Gender, male (female)	78 (19)
Ethnicity, number(%)	
Caucasian	67 (69.1)
Afro-American	9 (9.3)
Latin-American	3 (3.1)
South-Asian	1 (1)
Others	17 (17.5)
Matrimonial status, number(%)	
Single	77 (79.4)
Divorced	8 (8.2)
Married	6 (6.2)
Other	5 (5.2)
Employment status, number(%)	
Currently employed	27 (27.8)
Unemployed	70 (72.2)
Highest schooling completed (years)	
8 or less	10 (10.3)
9 to 10	16 (16.5)
11 or more	71 (73.2)

Table 3. Features associated with a positive prediction of violent gestures

Features	**Item of the questionnaire**	**Coefficients**
Stimulant use (past 12 months) with withdrawal symptoms	SCID Stimulants Item 5	37,9746297
Cannabis use (past 12 months) after having decided not to	CUPIT	22,6386

	Question 16	51
Cannabis use (past 12 months) and difficulties to go through a day without consuming	CUPIT Question 10	17,4570396
Cocaine use (lifetime) and relation between use and interpersonal problems	SCID Cocaine Item 3	14,4068665
Cannabis use (past 12 months) and finding problematic the time spent consuming	SCID Cannabis Item 12	12,5384905
Stole while threatening someone prior to the age of 15	SCID ASPD Item 128	10,8804552
Frequent temper tantrums between ages of 16 and 17	BGH Question 3c	9,38668745
Voluntarily destroyed other's belongings prior to the age of 15	SCID ASPD Item 131	8,58633195
Cannabis use (lifetime) and found the quantity consumed is larger than desired	SCID Cannabis Item 6	6,98749508
History of temper tantrums after the age of 18	BGH Question 3d	6,08869175
Prior episode of major depression	SCID Depression	6,04501004
Stimulant binges (past 12 months)	SCID Stimulants	6,02627434
Cocaine use (lifetime) and had difficulties fulfilling their obligations	SCID	5,43537

	Cocaine Item 1	226
When upset, becomes mistrustful against others or become confused and disoriented	SCID BPD Item 120	4,97935736
Conducted an illegal act prior to the age of 12 and did not get caught	BGH Question 8a	4,87430774
Somatic concerns	PANSS Question G1	4,23112942
Threatened or hurt someone with a weapon, a bat, a brick, a broken bottle, a knife or a gun prior to the age of 15	SCID ASPD Item 125	4,22719523
Difficulties in maintaining a conversation in French or English (auto perceived)	Sociodemographic	3,312844
Frequently initiated fights prior to the age of 15	SCID ASPD 124	3,30111659
Alcohol use (lifetime) and experienced cravings	SCID Alcohol Item 11	3,24208496
Alcohol binges (past 12 months)	SCID Alcohol	2,79669749
History of fighting against another person after the age of 18	BGH Question 6d	2,73196634
Prior manic episodes	SCID Mania	1,21395841
Cocaine use (lifetime) and saw a decrease in their social activities	SCID Cocaine Item 9	0,61978461

Alcohol binges (lifetime)	SCID Alcohol	0,55738073
Alcohol use (past 12 months) and found it difficult to control their consumption	SCID Alcohol Item 7	0,44206005

Table 4. Features linked with negative prediction of violent gestures

Features	Item of the questionnaires	Coefficients
Cocaine use (lifetime) and experienced withdrawal symptoms	SCID_TLU_Cocaine_Lifetime_5	-0,27987877
Cannabis use every day (last 3 months)	CUPIT_Q2	-0,31709792
Never experienced disciplinary issues at school	BGH_Q1	-0,31878408
Experienced a major family conflict after the age of 18	BGH_Q5d	-0,3438045
Never experienced a problem with an officer of the law	BGH_Q9	-0,50088409
Cannabis use (lifetime) and having a psychological or medical problem	SCID_TLU_Cannabis_Lifetime_10	-0,72965874
Cannabis use (lifetime) daily in the morning	CUPIT_Q4	-0,80285525
Cannabis use (lifetime) and relation between use and interpersonal problems	SCID_TLU_Cannabis_Lifetime_3	-0,8492082
Never lied a lot before the age of 15 to obtain something	SCID_TPA_133	-0,98553276
Stimulants use (lifetime) and relation between use and interpersonal problems	SCID_TLU_Stimulants_Lifetime_3	-1,72044576

Alcohol use (lifetime) and consumed despite imminent danger	SCID_TLU_Alcool_Lifetime_2	-1,77158892
Cannabis use (past 12 months) despite having a psychological or medical problem	SCID_TLU_Cannabis_12mois_10	-2,46828238
Cocaine use (lifetime) and experienced cravings	SCID_TLU_Cocaine_Lifetime_11	-3,37028849
Never did something illegal and get caught	BGH_Q8	-3,38271317
Stimulants use (lifetime) consumed despite imminent danger	SCID_TLU_Stimulants_Lifetime_2	-4,01563569
Alcohol use (past 12 months) despite having a psychological or medical problem	SCID_TLU_Alcool_12mois_10	-5,03669695
Alcohol use (lifetime)and found problematic the time spent consuming	SCID_TLU_Alcool_Lifetime_8	-8,2677788
Cannabis use (past 12 months)and never spent a day without using	CUPIT_Q7	-9,70203965
Alcohol use (lifetime) and consumed despite imminent danger	SCID_TLU_Alcool_12mois_2	-11,2663869
Cannabis use (lifetime) consumed despite imminent danger	SCID_TLU_Cannabis_Lifetime_2	-17,1858466

While the profile of associations is mostly understandable (mostly aggressive behaviour and externalizing disorder profile), many association seem to be counterintuitive. This is an indication that this host of factor might reflect noise.

Article 2. Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms

Alexandre Hudon

Mélissa Beaudoin

Kingsada Phraxayavong

Laura Dellazizzo

Stéphane Potvin

Alexandre Dumais

Publié dans *JMIR Mental Health (2021).*
Article révisé par les pairs.

Abstract

Background:

A growing body of literature has detailed the use of qualitative analyses to measure the therapeutic processes and intrinsic effectiveness of psychotherapies, which yield small databases. Nonetheless, these approaches have several limitations and machine learning algorithms are needed.

Objective:

The objective of this study is to conduct a systematic review of the use of machine learning for automated text classification for small data sets in the fields of psychiatry, psychology, and social sciences. This review will identify available algorithms and assess if automated classification of textual entities is comparable to the classification done by human evaluators.

Methods:

A systematic search was performed in the electronic databases of Medline, Web of Science, PsycNet (PsycINFO), and Google Scholar from their inception dates to 2021. The fields of psychiatry, psychology, and social sciences were selected as they include a vast array of textual entities in the domain of mental health that can be reviewed. Additional records identified through cross-referencing were used to find other studies.

Results:

This literature search identified 5442 articles that were eligible for our study after the removal of duplicates. Following abstract screening, 114 full articles were assessed in their entirety, of which 107 were excluded. The remaining 7 studies were analyzed. Classification algorithms such as naive Bayes, decision tree, and support vector machine classifiers were identified. Support vector machine is the most used algorithm and best performing as per the identified articles. Prediction classification scores for the identified algorithms ranged from 53%-91% for the classification of textual entities in 4-7 categories. In addition, 3 of the 7 studies reported an interjudge agreement statistic; these were consistent with agreement statistics for text classification done by human evaluators.

Conclusions:

A systematic review of available machine learning algorithms for automated text classification for small data sets in several fields (psychiatry, psychology, and social sciences) was conducted. We compared automated classification with classification done by human evaluators. Our results show that it is possible to automatically classify textual entities of a transcript based solely on small databases. Future studies are nevertheless needed to assess whether such algorithms can be implemented in the context of psychotherapies.

Keywords

psychotherapy ; artificial intelligence ; automated text classification ; machine learning ; systematic review

Introduction

The intrinsic effectiveness of psychotherapies is generally measured through semistructured interviews or self-reported questionnaires [1-3]. However, these instruments have limitations in relation to constructs that can be set a priori, for which there are standardized measures available. To assess the intrinsic effectiveness of psychotherapies (the psychotherapeutic process itself), an increasing number of research teams have started to use qualitative methods. Although these approaches have inherent biases (e.g., data analysis subjectivity), mathematical algorithms can be used to reduce such biases. Furthermore, assessment of a psychotherapy's intrinsic effectiveness usually refers to an assessment of a patient's characteristics and the therapeutic process [4]. Studies often use therapy session transcripts to qualitatively evaluate psychotherapies [5]. For in-person therapies, transcriptions are often time-consuming and classifying therapeutic interactions under various themes (labels) for analysis is even more demanding. Machine learning is a potential solution to reduce the amount of labor-intensive work required [6]. With the increasing development of new psychotherapies for various psychopathologies, there is a higher need for tools to measure and understand their effectiveness.

Text mining is one of the few techniques used in psychiatry to derive data from the large number of interactions that occur during therapy sessions [7]. One such technique is the use of artificial intelligence by means of machine learning. It is currently being used in many areas in the medical field, ranging from surgical procedure analyses to medical diagnostics [8]. When attempting to classify textual entities from medical fields into various categories, the text is often classified into a few categories. This can be done by applying a set of rules to an algorithm to be used for classification and is usually facilitated by the nature of the entity being classified (eg, signs and symptoms relating to a particular diagnosis or treatment) [9]. Classification of therapeutic interactions can be tricky considering the vast array of information associated with the therapy itself, the ability of the patient to communicate, and the context in which the therapy is being conducted [10]. This leads to transcripts that may vary widely from patient to patient; therefore, the information is less directly interpretable than medical records or results. In relevant fields where such data is usually used for research, such as psychiatry and psychology, the use of machine learning in the context of text mining in psychotherapy has been limited [11]. Many algorithms are readily available to conduct automated text classification [12]. Simple probabilistic mathematical algorithms (ie, naive Bayesian probability algorithms) as well as more complex ones (ie, neural networks) are available via open access libraries on the web [13]. Machine learning algorithms often need large databases to adequately classify new data by creating training sets and testing sets [14-16]. Large databases, such as some seen in the field of internet-enabled cognitive behavioral therapy, are required for complex machine learning algorithms to adequately learn and classify new information [1]. However, in-person therapies often yield databases that are smaller than the ones generated by internet-enabled cognitive behavioral therapy because of the need for human-driven transcriptions. This creates a need to find potential algorithms that can operate on small databases [17,18]. A machine learning algorithm applicable for small databases is therefore needed for such cases.

The objective of this study is to conduct a systematic review of the use of machine learning for automated text classification for small databases in the fields of psychiatry, psychology, and social sciences to determine the best algorithm for automatically classifying the content of psychotherapy transcripts. This would provide an interesting solution for automated therapy annotations in the context of qualitative analysis and could generate data to enable the evaluation of therapeutic processes.

Methods

Search Strategies

A systematic search was performed in the electronic databases of Medline, Web Of Science, PsycNet (PsycINFO), and Google Scholar from their inception dates until 2021 using text words and indexing (MeSH) terms with keywords that were inclusive for the fields of psychiatry (eg, psychiatric, psychiatry), psychology (e.g., psychology, psychotherapy, neuropsychology) and social sciences (e.g., social science) and machine learning. Additional records identified through cross-referencing were used to find other studies. The fields of psychiatry, psychology, and social sciences were selected as they include a vast array of textual entities in the domain of mental health that can be reviewed. A complete electronic search strategy is available in

Multimedia Appendix 1. The search methodology was developed by the corresponding author and a librarian specialized in mental health at the Institut universitaire en santé mentale de Montréal. Searches were completed by AH and cross-validated by MB in May 2021. No setting, date, or geographical restrictions were applied. Searches were limited to English- or French-language sources.

Study Eligibility

Studies were included if they met the following criteria: (1) classification in various data categories of textual entities (e.g., medical records, letters, transcripts); (2) the study was conducted in the fields of psychiatry, psychology, or social sciences; (3) automated classification of text was conducted in more than 2 data categories (text was classified in more than two features); (4) automated text classification was conducted by machine learning (either supervised or unsupervised algorithms); and (5) the number of elements in the database used was less than 10,000, which corresponds to a small database. Although there is no consensus on what a small database is, we defined a small database as one that had a maximum of 10,000 items since 5000-10,000 items have been referred to as small samples in prior studies [19-21]. Studies that use a combination of many algorithms, instead of a single algorithm, were also included. Unpublished literature was excluded as well as studies using artificial intelligence algorithms outside the scope of machine learning.

Data extraction

Data were extracted with a standardized form and cross-verified for consistency and integrity by two authors, AH and MB. Information such as size of the database, number of classification categories, algorithms used, prediction success rate (in %), and interjudge agreement were recorded.

Results

Description of studies

Our systematic review assessed studies that used machine learning to classify text in the fields of psychiatry, psychology, and social sciences. This literature search identified 5442 articles that were eligible for our study after the removal of duplicates. Following abstract screening, 114 full articles were assessed in their entirety, of which 107 were excluded. The remaining 7 studies were analyzed. The flowchart for the inclusion of studies in this systematic review is found in Figure 1. The details of the studies are provided in

Multimedia Appendix 2. Notably, a limited number of articles on automated text classification with small databases were found. Studies that met inclusion criteria reported different types of documents used for automated annotation. Social medical content, such as forum posts in the study by Yu et al [22] and Twitter entries in the study by Balakrishnan et al [23] generated the largest data sets (5000 and 5453 items, respectively). Those textual entities consisted of complete or partial sentences manually written by users and were annotated in their entirety. The remaining types of documents were mainly medical records completed by physicians or health science professionals. No image or mathematical data were classified by the algorithms as part of these studies.

-- Please insert Figure 1 here –

Algorithms

Overview

Several algorithms have been used on the presented textual entities. Naive Bayes classifier, decision tree–based algorithms, support vector machine (SVM) classifiers, and combinations of multiple algorithms were the main strategies used by the included studies. The number of categories for text classification ranged from 4-7 and overall precision classification ranged from 77.0%-91.8%. For the studies that included multiple algorithms, SVM-based algorithms demonstrated the best accuracy in 5 of 7 studies.

Naive Bayes Classifier

A naive Bayes classifier is a probabilistic-based classifier that makes use of Bayes' theorem to classify items into different categories [12]. This type of classifier achieves average performance in the context of supervised learning [24]. This type of algorithm is advantageous when little data is available as it can be optimally parameterized in the event of a small data set [25]. This algorithm assumes that there is independence between the predictors. For text classification, Balakrishnan et al [23] outlined that this algorithm works best when using each word as a variable that needs to be classified.

Decision Tree–Based Classifiers

Decision tree–based classifiers are nonparameterized; they are supervised learning methods that can be used to classify items [26]. Observations about an item are represented as branches and conclusions about an item's value (score) are represented as leaves [27]. Splitting across the different branches is based on defined rules according to the categories used to classify the items. In text classification, the general idea is that every piece of text being classified is split across the branches until it reaches a leaf (category) based on probabilistic rules set by the designer of the tree [27].

SVM Classifiers

SVM classifiers can be used in both supervised and unsupervised learning contexts. In simple terms, these classifiers use the concept of a hyperplane that divides a data set into classes. A hyperplane in an n-dimensional Euclidean space is a flat, n–1 dimensional subset of that space that divides the space into two disconnected parts [28]. The items in the data set are considered as data points on the hyperplane. The item being classified is therefore categorized in one of the disconnected parts.

Outcomes

In the 7 identified studies, SVM classifiers and algorithms combined with SVM classifiers tended to achieve the best prediction score (in %) as compared to other algorithms for small data sets. Studies by Zolnoori et al [29], Singh et al [30], and Yu et al [22] reported prediction scores of SVM classifiers that were superior to other classifiers for their data sets. Their precision scores ranged from 77%-90%. Only 3 studies attempted to compare the classification done by the classifiers with human annotators. The statistics used to assess these automated annotations were κ and pairwise agreements. The interrater agreement of these studies was comparable to interrater agreements for annotation done by human annotators; the κ scores were 0.84 [23], 0.67 [30], and 0.86 [29], respectively.

Discussion

Review of findings

In this study, we conducted a systematic review to identify potential algorithms that could be useful for small databases for the automatic annotation of unannotated interview transcripts from the field of psychotherapy. The systematic review we conducted demonstrated that limited literature exists on the subject. However, few algorithms displayed sufficient accuracy when performing text classification on small databases. SVM classifiers tended to display the best accuracy in the context of small databases.

Compared to other reviews on the subject, this study highlights algorithms being used in the context of small data sets, which is consistent with the reality of studies of therapies [31], as transcribing therapy sessions is time-consuming and demanding. Regarding novel therapy developments, such as virtual reality–based therapy, this is even more needed considering the small number of patients that have received these treatments so far [32]. Therapy usually involves a wider range of words and contextual sentences compared to other areas of medicine where specific words (e.g., symptoms, signs) can be used to facilitate classification. Therefore, it is not surprising to see that this systematic review identified algorithms that differ from those that are widely used in other medical fields. For example, Srivastava et al [33] reviewed the efficiency of different text classifiers in the context of social media posts referring to medical content. They found that a multilayer perceptron–based neural network performed best in their study as compared to a SVM classifier. Another study, conducted by Visveswaran and colleagues [34], identified convolutional long short-term memory neural networks as the best at predicting vaping habits. This can be explained by the fact that most classifiers are combined with a vectorizer when used to classify textual entities. A vectorizer transforms text into a meaningful number vector that can then be used by classifiers [35]. Considering that classification of textual entities to identify a specific diagnosis or medical condition usually requires specific terms that pertain to the diagnosis or condition, vectors tend to discriminate better between the textual entities of these fields [36]. This is usually not the case with therapy transcripts in the context of analysis of the psychotherapeutic process as this analysis often requires a larger array of categories that can sometime overlap.

In contrast with other types of medical data—such as imagery or numerical entities (eg, laboratory results)—where neural networks seem to be the most used class of algorithms for classification, textual classification appears to be performed with a more restricted number of classifiers [37]. This can be explained by the fact that text classification requires additional considerations. Automated classifications lack the ability to interpret a sentence out of a given context (eg, a therapeutic session), while the meaning of a sentence could change based on the context. Another complexity is that words can refer to different entities based on the sociocultural context. Therefore, considering such complexities can require further parameterizations and considerations, which may also explain why, in the identified studies, the same algorithm used on data sets of a similar size could have a diverging predictive score.

Consistent with our findings, linear SVM classifiers tend to be regarded as one of the best text classifying algorithms in the literature [38]. Many types of classifiers are available, but it appears that only a few are consistently used for the classification of textual entities [26]. This is consistent with our review, as the identified studies tended to use similar strategies when classifying textual entities. A recent literature review on data classification of clinical text data explains this phenomenon by the fact that there is a bottleneck of annotations in the context of supervised learning [39].

Limitations

This systematic review of literature focuses on the fields of psychiatry, psychology, and social sciences to reflect the type of textual entities usually found in therapy transcripts. A limitation of this study is the small number of classification algorithm studies published in these fields. As this is an emerging domain, the number of studies on the topic should increase in the future.

Conclusions

Machine learning can be beneficial for the field of psychiatry. Automated text classification for psychotherapy is a promising avenue to generate quantitative and qualitative data in an efficient way to make the data readily available for analyses. SVM classifiers appear to be preferred over other types of classifiers in the context of small databases. Using such classifiers could be useful in the evaluation of therapeutic processes of novel therapies where data are limited. Nevertheless, the limited number of articles found on the subject outlines the need for more development in this field, especially regarding the use of such classifiers in the domain of mental health.

Acknowledgements

This study was funded by Le Fonds de recherche du Québec – Santé (FRQS) and Services et recherches psychiatriques AD.

Authors' contributions

The study was designed by AH, SP, and AD. Statistical analyses were performed by AH and MB. All the authors have made substantial contributions and have revised, edited, and approved the manuscript.

Conflicts of Interest

None declared.

References

1. Ewbank, M.P., et al., Quantifying the Association Between Psychotherapy Content and Clinical Outcomes Using Deep Learning. JAMA Psychiatry, 2019.

2. Cook, S.C., A.C. Schwartz, and N.J. Kaslow, Evidence-Based Psychotherapy: Advantages and Challenges. Neurotherapeutics, 2017. 14(3): p. 537-545.

3. Hill, C.E., H. Chui, and E. Baumann, Revisiting and reenvisioning the outcome problem in psychotherapy: An argument to include individualized and qualitative measurement. Methodological issues and strategies in clinical research, 4th ed. 2016, Washington, DC, US: American Psychological Association. 373-386.

4. Szymanska, A., K. Dobrenko, and L. Grzesiuk, Characteristics and experience of the patient in psychotherapy and the psychotherapy's effectiveness. A structural approach. Psychiatr Pol, 2017. 51(4): p. 619-631.

5. Perepletchikova, F., On the Topic of Treatment Integrity. Clin Psychol (New York), 2011. 18(2): p. 148-153.

6. Sebastiani, F., Machine Learning in Automated Text Categorization. ACM Computing Surveys, 2002. 34(1): p. 1-47.

7. Abbe, A., et al., Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatr Res, 2016. 25(2): p. 86-100.

8. Khalid, S., et al., Evaluation of Deep Learning Models for Identifying Surgical Actions and Measuring Performance. JAMA Netw Open, 2020. 3(3): p. e201664.

9. Durstewitz, D., G. Koppe, and A. Meyer-Lindenberg, Deep neural networks in psychiatry. Mol Psychiatry, 2019. 24(11): p. 1583-1598.

10. Deo, R.C., Machine Learning in Medicine. Circulation, 2015. 132(20): p. 1920-30.

11. Cao, H., A. Meyer-Lindenberg, and E. Schwarz, Comparative Evaluation of Machine Learning Strategies for Analyzing Big Data in Psychiatry. Int J Mol Sci, 2018. 19(11).

12. Kowsari, et al., Text Classification Algorithms: A Survey. Information, 2019. 10(4).

13. Hämäläinen, W. and M. Vinni. Comparison of Machine Learning Methods for Intelligent Tutoring Systems. 2006. Berlin, Heidelberg: Springer Berlin Heidelberg.

14. Wanigasekara, C., et al., Improved Learning from Small Data Sets Through Effective Combination of Machine Learning Tools with VSG Techniques, in 2018 International Joint Conference on Neural Networks (IJCNN). 2018, IEEE.

15. Craig, T.K.J., et al., AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. The Lancet Psychiatry, 2018. 5(1): p. 31-40.

16. Dellazizzo, L., et al., Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychother, 2018. 25(6): p. 878-885.

17. Leff, J., et al., Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. Br J Psychiatry, 2013. 202: p. 428-33.

18. Leff, J., et al., Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? Psychosis, 2014. 6(2): p. 166-176.

19. du Sert, O.P., et al., Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophr Res, 2018. 197: p. 176-181.

20. Beaudoin M, P.S., Machalani A, Dellazizzo L, Bourguignon L, Phraxayavong K, Dumais A, The therapeutic processes of Avatar Therapy: A content analysis of the dialogue between treatment resistant patients with schizophrenia and their avatar. Psychology and psychotherapy: Theory, Research and Practice, 2020.

21. Shiner, B., et al., Automated classification of psychotherapy note text: implications for quality assessment in PTSD care. J Eval Clin Pract, 2012. 18(3): p. 698-701.

22. Slonim, N. and N. Tishby. The power of word clusters for text classification. in 23rd European Colloquium on Information Retrieval Research. 2001.

23. Joachims, T. Transductive inference for text classification using support vector machines. in Icml. 1999.

24. QDA Miner (Version 5). (2016): Provalis Research.

25. Ozaydin, B., et al., Text-mining analysis of mHealth research. Mhealth, 2017. 3: p. 53.

26. Shridhar, K., Subword Semantic Hashing for Intent Classification on Small Datasets. International Joint Conference on Neural Networks (IJCNN), 2019: p. 1-6.

27. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

28. Oliphant, T.E., Python for Scientific Computing. Computing in Science & Engineering, 2007. 9(3): p. 10-20.

29. Lazaro S.P. Busagala, W.O., Tetsushi Wakabayashi, Fumitaka Kimura, Multiple Feature-Classifier Combination in Automated Text Classification. IEEE Computer Society, 2012. 10th IAPR International Workshop on Document Analysis Systems.

30. Veronese, E., et al., Machine learning approaches: from theory to application in schizophrenia. Comput Math Methods Med, 2013. 2013: p. 867924.

31. Afshin Gholamy, V.K., Olga Kosheleva, Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. Departmental Technical Reports (CS), 2018.

32. Dell Zhang, J.W., Xiaoxue Zhao Estimating the Uncertainty of Average F1 Scores. ICTIR '15:

Proceedings of the 2015 International Conference on The Theory of Information, 2015: p. 317-320.

33. Allen, M., Intercoder Reliability Techniques: Scott's Pi. The SAGE Encyclopedia of Communication Research Methods, 2018: p. 753-755.

34. Balakrishnan, V., S. Khan, and H.R. Arabnia, Improving cyberbullying detection using Twitter users' psychological features and machine learning. Computers & Security, 2020. 90: p. 11.

35. Zolnoori, M., et al., A systematic approach for developing a corpus of patient reported adverse drug events: A case study for SSRI and SNRI medications. J Biomed Inform, 2019. 90: p. 103091.

36. Karystianis, G., et al., Automatic mining of symptom severity from psychiatric evaluation notes. International Journal of Methods in Psychiatric Research, 2018. 27(1): p. 11.

37. Singh, V.K., et al., Machine learning for psychiatric patient triaging: an investigation of cascading classifiers. J Am Med Inform Assoc, 2018. 25(11): p. 1481-1487.

38. Clark, C., et al., Automatic classification of RDoC positive valence severity with a neural network. J Biomed Inform, 2017. 75s: p. S120-s128.

39. Dai, H.J. and J. Jonnagaddala, Assessing the severity of positive valence symptoms in initial psychiatric evaluation records: Should we use convolutional neural networks? Plos One, 2018. 13(10).

40. Yu, L.C., et al., Mining association language patterns using a distributional semantic model for negative life event classification. J Biomed Inform, 2011. 44(4): p. 509-18.

41. Haddoud, M., et al., Combining supervised term-weighting metrics for SVM text classification with extended term representation. Knowledge and Information Systems, 2016. 49(3): p. 909-931.

42. Han Liu, M.C., Semi-random partitioning of data into training and test sets in granular computing context. Granular Computing, 2017. 2: p. 357-386.

43. Balakrishnan, V., S. Khan, and H.R. Arabnia, Improving cyberbullying detection using Twitter users' psychological features and machine learning. Computers & Security, 2020. 90.

44. Singh, V.K., et al., Machine learning for psychiatric patient triaging: an investigation of

cascading classifiers. Journal of the American Medical Informatics Association, 2018. 25(11): p. 1481-1487.

45. de Ávila Berni, G., et al., Potential use of text classification tools as signatures of suicidal behavior: A proof-of-concept study using Virginia Woolf's personal writings. PLoS One, 2018. 13(10): p. e0204820.

46. Mak, K.K., K. Lee, and C. Park, Applications of machine learning in addiction studies: A systematic review. Psychiatry Res, 2019. 275: p. 53-60.

47. Venkatasubramanian, S., et al. A non-syntactic approach for text sentiment classification with stopwords. in Proceedings of the 20th international conference companion on World wide web. 2011.

Abbreviations

SVM: support vector machine

Figures and Tables

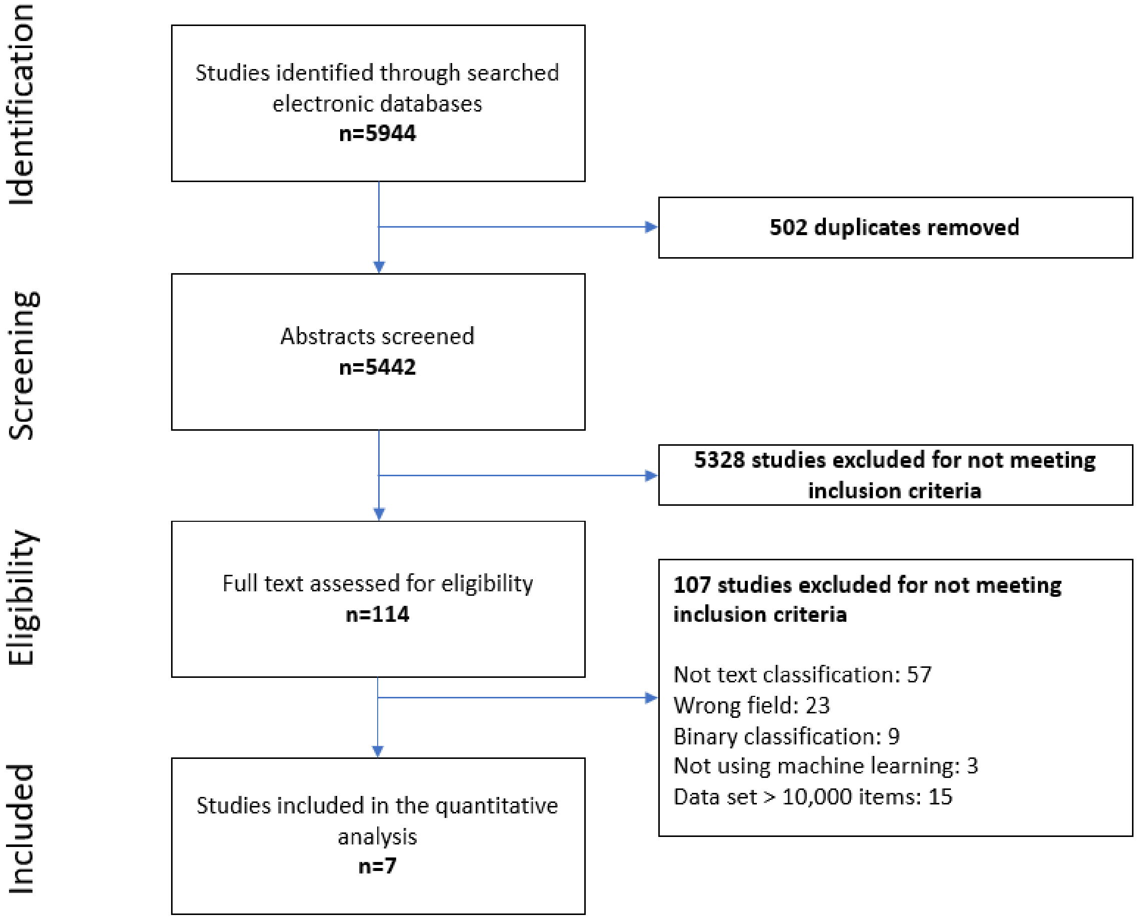
Identification
Screening
Eligibility
Included
Studies identified through searched electronic databases
n=5944
502 duplicates removed
Abstracts screened
n=5442
5328 studies excluded for not meeting inclusion criteria
Full text assessed for eligibility
n=114
107 studies excluded for not meeting inclusion criteria
Not text classification: 57
Wrong field: 23
Binary classification: 9
Not using machine learning: 3
Data set > 10,000 items: 15
Studies included in the quantitative analysis
n=7

Figure 1. Flowchart depicting the process of study selection.

Table 1. Multimedia Appendix 1 - Electronic search strategy for the systematic review conducted.

Database; Search	Search Terms
PubMed; k= 2868	("Machine Learning"[Mesh] OR "Natural Language Processing"[Mesh] OR "Data Mining"[Mesh] OR "Machine learning"[TIAB] OR "deep learning"[TIAB] OR "text mining"[TIAB] OR "data mining"[TIAB] OR "learning algorithm"[TIAB] OR "learning algorithms"[TIAB] OR "classification algorithm"[TIAB] OR "classification algorithms"[TIAB] OR "language processing"[TIAB] OR "text analysis"[TIAB]) AND Psychiatric[TIAB] OR Psychiatry[TIAB] OR Psychotherapy[TIAB] OR psychotherapies[TIAB] OR therapy[TIAB] OR therapies[TIAB] OR Psychology[TIAB] OR Neuropsychology[TIAB] OR Psychological[TIAB] OR neuropsychological[TIAB] OR "Social science"[TIAB] OR "social sciences"[TIAB])
Web of Science; k= 2736	(Psychiatric OR Psychiatry OR Psychotherapy OR Psychotherapies OR Therapy OR Therapies OR Psychology OR Neuropsychology OR Psychological OR Neuropsychological OR Social science OR Social sciences) AND TS= (Machine learning OR Deep learning OR text mining OR data mining OR learning algorithm OR learning algorithms OR classification algorithm OR classification algorithms OR language processing OR text analysis)
PsychInfo; k = 4	exp machine learning/ OR exp data mining/ OR exp automated information processing/ OR ("Machine learning" or "deep learning" or "text mining" or "data mining" or "learning algorithm" or "learning algorithms" or "classification algorithm" or "classification algorithms" or "language processing" or "text analysis").ab. or ("Machine learning" or "deep learning" or "text mining" or "data mining" or "learning algorithm" or "learning algorithms" or "classification algorithm" or "classification algorithms" or "language processing" or "text analysis").ti. AND (Psychiatric or Psychiatry or Psychotherapy or psychotherapies or Therapy or therapies or Psychology or

	Neuropsychology or Psychological or neuropsychological or "Social science" or "social sciences").ab. or (Psychiatric or Psychiatry or Psychotherapy or psychotherapies or Therapy or therapies or Psychology or Neuropsychology or Psychological or neuropsychological or "Social science" or "social sciences").ti.
Google Scholar; k= 336	(allintitle:"Machine learning" OR "Text mining" OR "classification algorithm" OR "language processing" OR "text analysis") AND ("Psychiatry" OR "Psychiatric" OR "Psychology" OR "Social sciences" OR "social sciences "OR "Neuropsychology" OR "therapy" OR "therapies" OR "psychotherapy" OR "psychotherapies") AND ("text classification" OR "text processing")

Table 2. Multimedia Appendix 2 - Detailed results of the systematic review study selection.

Studies	Dataset size	Datatype	# categories	Algorithms used	Algorithm with highest accuracy	Precision	Recall	Prediction Statistic	Prediction Score (%)	Inter-judge agreement statistic	Inter-judge agreement statistic
[22] Balakrishnan V, Khan S, Arabnia H.R., 2020	5453	Twitter entries	4	NB, RF, J48	J48 (DT based)[a]	N/A	0.97	F1-Score	91.88	Kappa	0.84
[23] Zolnoori et al., 2019	891	drug review posts	6	SVM , RF, NB, LR	SVM	0.91	0.912	F1-Score	90.06	Pairwise agreement	0.86
[24] Karystianis et al, 2018	541	psychiatric records	4	Rule-based method, NN	Rule-based method	N/A	N/A	MAE	80.1	N/A	N/A
[25] Singh et al., 2018	325	Initial psychiatric assessment record	7	LSVM and other algorithms	One-class-at-a-time (LSVM and others) [b]	N/A	N/A	INMAE	77	Kappa	0.67
[26] Clark et al., 2017	600	Patient clinical notes	4	MLP and MLR classifiers	MLP with a 20% drop-rate N/A	N/A	N/A	MAE	77.86	N/A	N/A
[27] Dai H.J., Jonnagaddala J, 2018	649	Patient records	4	CNN, SVM, C4,5, NB	CNN overall SVM for 2 categories	N/A	N/A	MAE	Mean: 53.9 Normalized :78.5	N/A	N/A

[28] Yu et al., 2011	5000	Forum posts	5	NB, C4,5 , TAN, SVM	SVM	N/A	N/A	F1-Score	79.5 to 81.9	N/A	N/A

Abbreviations: NB = Naïve Bayes, LR = Logistic Regression RF= Random forest, DT= Decision Tree, SVM = Support vector machine, NN = Neural Network, LSVM = Linear support vector machine, MLP = Multiple layered platform, CNN= Convolutional Neural Network, TAN= Tree Augmented Naïve Bayes. aJ48 is a decision-tree based algorithm often combined with an SVM. b In the One-class-at-a-time cascading algorithm, linear SVM was the best performing algorithm with an accuracy of 61%.

Article 3. Deciphering the Mosaic of Therapeutic Potential: A Scoping Review of Neural Network Applications in Psychotherapy Enhancements

Alexandre Hudon

Maxine Aird

Noémie La Haye-Caty

Publié dans *BioMedInformatics (2023).*
Article révisé par les pairs.

Abstract

Background: Psychotherapy is a component of the therapeutic options accessible in mental health. Along with psychotherapy techniques and indications, there is a body of studies on what are known as psychotherapy’s common factors. However, up to 40% of patients do not respond to therapy. Artificial intelligence approaches are hoped to enhance this and with the growing body of evidence of the use of neural networks (NNs) in other areas of medicine, this domain is lacking in the field of psychotherapy. This study aims to identify the different uses of NNs in the field of psychotherapy. Methods: A scoping review was conducted in the electronic databases EMBASE, MEDLINE, APA, and CINAHL. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement influenced this study’s design. Studies were included if they applied a neural network algorithm in the context of a psychotherapeutic approach. Results: A total of 157 studies were screened for eligibility, of which 32 were fully assessed. Finally, eight articles were analyzed, and three uses were identified: predicting the therapeutic outcomes, content analysis, and automated categorization of psychotherapeutic interactions. Conclusions: Uses of NNs were identified with limited evidence of their effects. The potential implications of these uses could assist the therapist in providing a more personalized therapeutic approach to their patients. Given the paucity of literature, this study provides a path for future research to better understand the efficacy of such uses.

Keywords

psychotherapy; neural networks; artificial intelligence; mental health; psychology; psychiatry

Introduction

Psychotherapeutic interventions

Psychotherapy is an important part of the array of treatments available to help patients suffering from a vast variety of mental illnesses. The evolution of systematized, evidence-based psychotherapeutic approaches in the field of mental illness can be traced back to the early 20th century [1]. Over the years, various psychotherapeutic methods have been developed and widely adopted. The American Psychological Association defines psychotherapy as a collaborative treatment based on the relationship between a person and a psychotherapist [2]. Over the 20th century, three major streams of psychotherapy have emerged. The first originates from Freud's pioneering work, which proposed a coherent approach emphasizing the major influence of the unconscious mind on our daily lives [3]. The second, rooted in scientific observation of human behavior, gave rise to the cognitive behavioral movement. The third revolves around humanistic approaches, prioritizing phenomenological perspectives and self-determination in treatment [4]. Numerous meta-analyses have consistently demonstrated the efficacy of psychotherapy in mental health disorders, in some cases showing comparable outcomes to pharmacological treatments [5]. Psychotherapeutic treatments can benefit individuals of all ages, educational levels, and ethnic and cultural backgrounds [6]. Given the wide range of efficacy of psychotherapy, the choice of psychotherapy type is generally guided by evidence-based validation for specific medical conditions [7]. For instance, cognitive-behavioral therapies, interpersonal therapy, and behavioral activation are considered first-line evidence-based acute and maintenance psychological treatments in depressive disorders [6]. Furthermore, meta-analyses suggest strong evidence for short- and long-term cognitive behavioral approaches to alleviate symptom distress in psychotic disorders [8].

The duration of psychotherapy can vary significantly [9,10,11]. Meta-analyses on different psychotherapeutic modalities indicate a significant improvement in symptoms in 53% of patients after eight weekly sessions, increasing to over 83% after 52 sessions [11]. There are no formal contraindications for psychotherapy [12]. However, therapeutic modalities must be carefully evaluated and adjusted to each patient's needs, as inappropriate psychotherapy, like other medical treatments, could have adverse effects [13].

Common Factors across Psychotherapeutic Approaches

Alongside the specific psychotherapy modalities and their indications, there exists an extended body of research on what are known as the common factors of psychotherapy. These are a set of characteristics present across all types of therapies that have been defined as early as 1936 and are considered fundamental to achieving positive psychotherapeutic outcomes. The factors that have been highlighted as most contributive to favorable outcomes are the therapeutic alliance, therapist empathy, goal consensus and collaboration, positive regard, mastery, genuineness, mentalization, emotional experience, and client expectations [14]. In a therapeutic dyad or a group setting, it has been found that relationship factors, some of them related to the common factors, are correlated with improved levels of functioning [15]. The outcomes of psychotherapy are conceptualized in a myriad of different ways but can be broadly described on a multidimensional level to include symptom reduction, improvement in functioning and quality of life, achievement of collaboratively articulated therapy goals, and a mature shift in defenses [14]. Moreover, it is important to understand that the common factors are not merely a set of elements that can be identified in all psychotherapies; rather, they 'collectively shape a theoretical model about the mechanisms of change in psychotherapy [16]. It is stipulated that benefits in psychotherapy are produced through three pathways, and these pathways have underlying mechanisms that stem from 'evolved characteristics of humans as social species' [16,17]. According to Wampold, the pathways are (1) the real relationship, (2) the creation of expectations through the explanation of disorder and treatment, and (3) the enactment of health-promoting actions [16].

The most extensively researched common factor is the therapeutic alliance, or the working relationship between patient and therapist. It is composed of the bond between patient and therapist as well as the agreement on therapeutic goals and the tasks of therapy. Its beneficial effect is therefore based on an increase in the mutuality and investment of the patient and the therapist in the therapy, as well as an increase in resilience and tolerance of distressing affects [14]. Another common factor shown to have a major curative effect on therapy is therapist empathy, which is both an inherent quality and learned skill. It is a complex phenomenon that can be subdivided into different factors that include mimicry, emotional and affective sharing, compassion, and sympathy. The therapist that achieves empathy is also able to distinguish the source of emotions in the therapeutic dyad, which refers to an awareness of the countertransference at play. This enables the therapist to adequately identify and reflect on the emotions expressed by the patient; this capacity, combined with the creation of a 'holding environment' in which emotions are validated and overwhelming affects are contained by the therapist, create beneficial outcomes by increasing ego strength. A focus on common factors can not only improve psychotherapy outcomes, but also facilitate an integration of common recommendations for effective psychotherapy training [14,18].

Artificial intelligence in the Field of Psychotherapy

The literature on artificial intelligence supports that there are several applications of specialized tools and techniques that could be employed to enhance psychotherapy [19]. As an example, it has been stated that up to 40% of patients do not respond to therapy, and that artificial intelligence is hoped to enhance this number by using close or real-time recommendations [20]. Furthermore, a recent study suggested that artificial intelligence will have a beneficial impact, but that further empirical analysis through data-driven model development is needed [21]. It has therefore been hinted that artificial intelligence, especially the use of deep learning models, might help in personalizing patient treatments [22].

One such approach is known as neural networks (NNs). Algorithms relating to NNs are made up of node layers, each of which has an input layer, one or more hidden layers, and an output layer [23,24]. Each node, or artificial neuron, is linked to another and has its own weight and threshold. If the output of any node exceeds the given threshold value, that node is activated and begins transferring data to the network's next tier [23]. Otherwise, no data are sent to the next network layer. The uses of NNs could provide potential enhancements to psychotherapy, but not such compilation of evidence exists to our knowledge in the current literature.

Objectives and hypotheses

The aim of this study is to identify the different uses of NNs in the field of psychotherapy. This will provide a first insight into therapeutic potential in psychotherapy enhancements with the help of this fast-growing and widely used modality that is currently being used in many other areas of medicine. We hypothesize that even though this field is emergent, there will be a wide array of uses such as predictive analysis, intervention delivery, and content analyses. This scoping review will provide a better overview on these different uses and provide details about key areas of future developments in the integration of NNs in the enhancement of psychotherapeutic approaches.

Methods

Search Strategies

From their inception dates through 2023, a systematic scoping search was conducted in the electronic databases EMBASE, MEDLINE, APA, and CINAHL by the authors. Search strategies included the use of text words and indexing terms (MeSH) containing keywords targeting the area of psychotherapy (e.g., psychotherapy, therapy, intervention, psychotherapeutic approaches, etc.) and the field of neural networks (e.g., machine learning, neural networks, artificial neural networks). These broad phrases were chosen because they better describe the use of neural networks in the context of psychotherapeutic approaches. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement influenced this study's design [25]. The Appendix A contains the main electronic search approach that was used. The search approach was conceptualized and implemented with the assistance of an experienced librarian specializing in mental health. There were no setting or geographical constraints imposed. The search approach used was limited to sources in English or French as to prevent interpretation biases.

Study Eligibility

Studies were considered if they met the following criteria: (1) the study is about (or uses) a type of neural network; (2) the study was conducted in the field of psychiatry or psychology and focuses specifically on mental health; (3) the neural network was used in the context of a psychotherapeutic approach; and (4) the study was focused on a clinical or future clinical intervention. Protocols and proof-of-concept research were excluded as exclusion criteria. There were no unpublished publications (or preprints) included in this review.

Data Extraction

Data were extracted using a standardized form (in Microsoft Excel, version Microsoft 365, EULA license, USA) and separately counter-verified for consistency and integrity by the authors (A.H., M.A.). NHC addressed any disagreements regarding the inclusion (or exclusion) of a study. The following information was systematically extracted: authors, population, type of neural networks, psychotherapeutic interventions and/or clinical application over a psychotherapeutic approach, psychometric tools (if used) and main outcomes identified.

Quality Assessment

Criteria derived from the GRADE Checklist were used to assess the quality of the identified studies. As per the GRADE methodology, studies were graded from very low, low, moderate-to-low, moderate, moderate-to-high, and high [26]. Evidence will be evaluated as weaker if it has risks of bias, if there are inconsistencies between the recommendations and the data presented, if it is declined indirectly from the results and if it is imprecise. On the contrary, it will be higher if the reported effect size is large, if there is an identified causality phenomenon, if the analysis of the results correctly identifies the confounding variables or suggests an absence of effect when it is the case. In principle, randomized controlled studies, according to this rating system, can achieve a maximum quality of high, while observational studies can have a quality of low at best. The GRADE system therefore uses eight criteria to critically appraise the quality of evidence: risk of bias or study limitations, inconsistency of results, indirectness of evidence, imprecision, reporting bias, effect size, dose–response gradient, and direction of plausible biases [26].

The author AH rated each studies using the GRADE Checklist. Quality assessment was then reviewed independently by MA to ensure consistent findings. Inconsistencies in the grading were discussed by the authors.

Results

Description of studies

This scoping review's search strategies enabled the retrieval of 430 potential studies. Across the different databases from which these studies were identified, 273 duplicates were removed with the assistance of the EndNote software (version 20, Clarivate). The remaining 157 studies were summarily screened for eligibility. Amongst these, 123 were excluded as they did not meet the inclusion criteria based on their title and abstract. From the remaining 34 studies, 32 were fully assessed as the full text of two studies could not be accessed. Of these 32 studies, 4 were excluded as they were not about the use of an NN, 7 because they were not in a psychotherapeutic context, 4 because they were not applicable to the field of mental health, and 9 because they were articles of the wrong type. Finally, eight studies were included in this scoping review, with various uses of NNs in the context of psychotherapeutic approaches for the fields of psychiatry, psychology, and mental health. The PRISMA flowchart for the inclusion of studies can be found in Figure 1. The studies identified and their details are presented in Table 1.

-- Please insert Figure 1--
-- Please insert Table 1—

Applications of Neural Networks in Psychotherapy

Predicting Patients' Psychotherapeutic Outcomes

One of the identified uses of NNs in the context of psychotherapy is in the prediction of patient's outcomes. A study by Gori and colleagues conducted on 150 Italian patients requesting psychotherapy used a basic ANN structure to predict patient's psychotherapeutic outcomes based on their MMPI-2 score [27]. The MMPI-2 is a personality questionnaire for diagnostic, descriptive and therapeutic purposes [35]. Each item is calculated using a T score. They used each patient’s MMPI-2 individual T scores and attempted to predict their clinical outcome as per the observed change in the T scores. They achieved a mean rate of 81% successful classification in the forecast of successful and unsuccessful treatment. The authors of this study state that such tool should not replace clinical judgement, and are designed to support the psychotherapist in their decision-making process. Such predictive tools could help the therapists in providing the best possible intervention for a specific patient. The authors of this study mention that the limited sample size can impact the result of the ANN and this should be carefully taken into consideration when deriving predictions.

In Koppe’s study of 2019, the authors conducted an analysis of different studies from the literature about the use of recurrent NNs in mobile sampling [29]. This is evaluated in the context of patients suffering from psychosis. They support that the use of such NNs could be used to forecast individual trajectories and schedule online feedback and interventions, and this should be further studied. As an example, a patient suffering from schizophrenia could have their trajectory predicted as per a mobile device, which is dynamically updated as time elapses, and different interventions could be conducted in response to the predicted trajectory to avoid a psychotic episode. Small datasets and varying data distributions are reported to be the main dangers of such NNs, and this should be considered when inferring a prediction.

As part of therapeutic outcomes, there is the notion of drop-out rates, which is of interest. The 2022 study by Bennemann and his colleagues compared the use of an ensemble of machine learning algorithms to a generalized linear model using neural network modeling [32]. With the data (per the Personality Style and Disorder Inventory and the Brief Symptom Inventory) of 2543 outpatients who were treated with CBT, they used several types of artificial NNs such as a feed-forward NN, an averaged feed-forward NN, and a monotone multi-layer perceptron NN, as well as other machine learning techniques. They demonstrated that for this dataset, NNs were identified to be less suited to predicting naturalistic datasets and binary events. Ensemble modeling comprising Random Forest algorithms and nearest-neighbor modelling performed best, by correctly identifying 63.4% of cases of patients who dropped out [32]. The quality of the data is reported to be poor, and might have hindered the results.

For patients suffering from tinnitus and following an internet-enabled cognitive behavioral therapy, artificial NNs and support vector machines were studied to predict their therapeutic outcomes. In a 2022 from Rodrigo and colleagues, 228 patients suffering from tinnitus who follow internet-enabled cognitive behavioral therapy had to complete the Tinnitus Functional Index (TFI) at the beginning of their treatment and at the end to determine if their therapy was successful (or not) [34]. The authors implemented two types of machine learning models: an artificial NN and a support vector machine, over 33 predictor variables including 7 demographic variables, 15 tinnitus and hearing-related variables, 4 treatment-related variables, 9 different types of tinnitus, and 7 clinical factors. The artificial NN performed best, with a predictive accuracy represented by an area under the curve of 0.73, as compared to the SVM, which achieved an area under the curve of 0.69 [34]. The authors state that his study was limited to the predictive variables, and other models could yield other results if encompassing other variables.

Content analysis

One application of NNs is content analysis. In Nitti et al. (2010) study, a discourse flow analysis was conducted over verbatim from psychodynamic therapy interventions [28]. In their work, they illustrate that the role of the discourse in psychotherapy is to generate new meanings through time, and this can be demonstrated by a discourse network. They created a discourse flow network and used it over the verbatim of a patient who went through 43 sessions of psychodynamic therapy. Using different metrics such as connectivity (density of the association among nodes), activity (global network ability to extend the spectrum of associations among nodes through time), and regulation (via super nodes, which are particular nodes carrying out the function of super-ordered meaning working as taken-for-granted assumptions), they concluded that NNs allow for the identification of patterns characterizing the psychotherapy process. It is mentioned that the redundancy of the language could potentially impact the validity of the model.

Another interesting approach was identified in Burger's study of 2021. They used though records and data from 320 healthy participants over depressive, anxious, and cognitive distortions scales to verify the possibility of using machine learning to extract schemas from their thought records [31]). Schemas are cognitive structures that make up our view of the world, and such schemas can be maladaptive, which is makes it interesting to be able to identify them. This is used in several therapeutic approaches. One of the machine learning approaches used is recurrent NNs, which were found to outperform the use of other algorithms such as k-nearest-neighbor and support vector machine for such a task. A larger set of schemas is suggested so that better content analysis can be achieved.

As part of psychotherapeutic content, emotions play an important role in the therapeutic relationship. A 2022 study by Chen and colleagues assessed the use of convolutional NNs for the recognition of emotions in students using human-interactive psychotherapy [33]. Their results demonstrate that deep learning convolutional NNs have better (accuracy of 81.86%) student emotion recognition ability than backpropagation neural networks (BPNNs) and decision tree algorithms (below 80%). They also conducted an acceptability analysis to identify the students' thoughts about emotional recognition and classification, conducted via artificial intelligence, and this methodology was found acceptable by the students. The limitations of the study are not mentioned in the manuscript.

Automated Categorization of Psychotherapeutic Interactions

A study conducted on the interactions of 13,073 patients demonstrated that recurrent NNa can perform well and achieve human-level agreement when categorizing therapist utterances [30]. This study was conducted with the use of a bidirectional long–short-term memory algorithm, which is a form of natural language processing using recurrent neural networks. It requires a large amount of data in order to achieve acceptable performances. Therefore, the authors used data from patients who followed internet-enabled CBT. Categorization of over 24 potential categories of utterances was achieved with acceptable performance (as characterized by precision) and recall of interactions classified in each of the categories [30]. The authors state that it is hard to know if a therapeutic intervention was conducted in the appropriate manner, and the model does not consider the quality of these interventions.

Quality of the Evidence

As can be observed, most studies achieved a low quality of evidence at best. Considering that most of the identified studies are conducted with a low sample size in the context of descriptive analysis, the quality of the evidence is therefore limited, per the GRADE system [26]. Further studies should be conducted on larger sample sets to confirm the effectiveness of the identified uses of NNs in the context of psychotherapy.

Discussion

Main Findings

The aim of this study was to identify the different use of NNs in the field of psychotherapy. A scoping review strategy was employed, considering the limited amount of literature on the subject. A total of eight articles were analyzed, and three main uses were identified: predicting therapeutic outcomes, content analysis, and automated categorization of psychotherapeutic interactions. While the number of studies identified were limited, this study provided a few examples of practical implication of NN for psychotherapeutic approaches and patient care.

In several areas of medicine, predictive approaches are being employed to select the best treatment for the patient. Prediction of patients' clinical trajectories, especially in the field of psychosis, is an eminent avenue of research in psychiatry. Such approaches could lead to more personalized treatments [36]. As identified in this scoping review, the use of NNs for predicting outcomes was brought up in Gori's study of 2010, which stated that such use of NNs could help the therapist with their decision making, and help in selecting the appropriate treatment for the patients [27]. However, the literature is currently hinting at the potential limitations of such uses, considering that models are dependent on the data they are provided; therefore, the transparency of such models should be made available to the clinicians to highlight their potential limitations [37]. Complex evaluations, such as posing a diagnosis, can cause predictive models to achieve poor performances, and can therefore limit their usability in assisting the clinician in their decision-making. The application of such models in the day-to-day work of psychologists and psychiatrists could also pose several ethical challenges, which should be addressed prior to broad usage [38]. In the field of psychotherapy, emerging randomized controlled trials using personalized prediction and adaptation tools for treatment outcomes are hoped to provide further information on the role of prediction tools in assisting therapists [39].

Machine learning has been used for content analysis in many domains of medicine [40]. For example, a recent literature review identified that text processing and text mining can be conducted using different kinds of artificial NN, recurrent NN, convolutional NN and linear short-term memory algorithms [41]. Different types of content can be analyzed in the context of psychotherapy, such as the content of the interactions, the emotions, and the relationship (therapeutic alliance). Content analysis in psychotherapy has been studied for several years, and the use of computational techniques to enhance this approach has been forecasted since over two decades ago [42]. However, it was found that limited data are available on the use of NNs for this context. While emotional recognition was found acceptable per Chen and colleagues, other work in the field of facial recognition demonstrates the superiority of NNs compared with other algorithms in recognizing facial traits [33,43].

In psychotherapy, implementation of supervised algorithms to categorize patient’s interactions have been employed, but are often limited by small datasets [44]. However, when such datasets are available, NNs offer acceptable performance, as was observed in Ewbank and colleagues’ study for a large array of categories [30]. When large datasets are not available, support vector machines are preferred, especially when the interactions are linearly separable [45]. One major concern for such applications, notably for NNs, is when data are imbalanced. Depending on the training data available, machine learning algorithms including NNs can perform poorly on some categories, which can lead to classification anomalies [46]. This limitation must be taken into account when conducting analyses on verbatims or other therapeutic corpora that have been automatically annotated.

Limitations

This scoping review focused only on the different uses of neural networks in the context of psychotherapy. The current search strategies highlight the use of NNs in the context of psychotherapy; however, certain studies might have used ensemble modeling comprising NNs. These might have been overlooked, considering it is difficult to estimate the role and effect of NNs when they are contained in an ensemble model. Furthermore, the small number of studies identified makes it difficult to derive significant conclusions as to the effects of the different uses of NNs in psychotherapy.

Conclusions

Psychotherapy is an important form of treatment for patients suffering from mental illness. To better understand psychotherapeutic processes and approaches, computational techniques have emerged across the years. This scoping review focused on the use of NNs in the context of psychotherapeutic approaches. From the eight studies analyzed, three main uses were identified: predicting therapeutic outcomes, content analysis, and automated categorization of psychotherapeutic interactions. The potential implications of these uses could assist the therapist in providing a more personalized therapeutic approach with their patients. Considering the limited amount of literature on the subject, this study paves the way for future research to better understand the effectiveness of such uses. Integration of NNs in the clinical aspects of psychotherapeutic approaches should be further studied, and their impact on clinical outcomes should also be studied.

Author Contributions

Conceptualization, A.H., M.A. and N.L.H.-C.; methodology, A.H., M.A. and N.L.H.-C.; validation A.H., M.A. and N.L.H.-C.; data curation, A.H., M.A. and N.L.H.-C.; writing—original draft

preparation, A.H., M.A. and N.L.H.-C.; writing—review and editing, A.H.; supervision, A.H.; project administration, A.H. All authors have read and agreed to the published version of the manuscript.

Fundings

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Marks S. Psychotherapy in historical perspective. Hist Human Sci. 2017;30(2):3-16.

2. Kazdin AE. Understanding how and why psychotherapy leads to change. Psychother Res. 2009;19(4-5):418-28.

3. Bargh JA, Morsella E. The Unconscious Mind. Perspectives on Psychological Science. 2008;3(1):73-9.

4. Solobutina MM, Miyassarova LR. Dynamics of Existential Personality Fulfillment in the

Course of Psychotherapy. Behav Sci (Basel). 2019;10(1).

5. Imel ZE, Wampold BE. The importance of treatment and the science of common factors in psychotherapy. Handbook of counseling psychology. 2008;4:249-66.

6. Kennedy SH, Lam RW, McIntyre RS, Tourjman SV, Bhat V, Blier P, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 3. Pharmacological Treatments. Can J Psychiatry. 2016;61(9):540-60.

7. Roth A, Fonagy P. What works for whom?: a critical review of psychotherapy research. 2006.

8. Lincoln TM, Pedersen A. An Overview of the Evidence for Psychological Interventions for Psychosis: Results From Me-ta-Analyses. Clinical Psychology in Europe. 2019;1(1):1-23.

9. Bergin AE, Garfield SL. Handbook of psychotherapy and behavior change. Handbook of psychotherapy and behavior change1994. p. 866 p.- p.

10. Garfield SL, Bergin AE, Dryden W. Handbook of psychotherapy and behavior change. Journal of Cognitive Psychotherapy. 1987;1(4):264-5.

11. Barkham M, Lambert MJ. The efficacy and effectiveness of psychological therapies. Bergin and Garfield's handbook of psychotherapy and behavior change: 50th anniversary edition, 7th ed. Hoboken, NJ, US: John Wiley & Sons, Inc.; 2021. p. 135-89.

12. Linden M, Schermuly-Haupt ML. Definition, assessment and rate of psychotherapy side effects. World Psychiatry. 2014;13(3):306-9.

13. Strauss B, Gawlytta R, Schleu A, Frenzl D. Negative effects of psychotherapy: estimating the prevalence in a random national sample. BJPsych Open. 2021;7(6).

14. Nahum D, Alfonso CA, Sönmez E. Common Factors in Psychotherapy. In: Javed A, Fountoulakis KN, editors. Advances in Psychiatry. Cham: Springer International Publishing; 2019. p. 471-81.

15. Fisher H, Atzil-Slonim D, Bar-Kalifa E, Rafaeli E, Peri T. Emotional experience and alliance contribute to therapeutic change in psychodynamic therapy. Psychotherapy (Chic). 2016;53(1):105-16.

16. Wampold BE. How important are the common factors in psychotherapy? An update.

World Psychiatry. 2015;14(3):270-7.

17. Hyland ME. A reformulated contextual model of psychotherapy for treating anxiety and depression. Clin Psychol Rev. 2020;80:101890.

18. Cook SC, Schwartz AC, Kaslow NJ. Evidence-Based Psychotherapy: Advantages and Challenges. Neurotherapeutics. 2017;14(3):537-45.

19. Tahan M. Artificial Intelligence applications and psychology: an overview. Neuropsychopharmacol Hung. 2019;21(3):119-26.

20. Gual-Montolio P, Jaén I, Martínez-Borba V, Castilla D, Suso-Ribera C. Using Artificial Intelligence to Enhance Ongoing Psychological Interventions for Emotional Problems in Real- or Close to Real-Time: A Systematic Review. Int J Environ Res Public Health. 2022;19(13).

21. Horn RL, Weisz JR. Can Artificial Intelligence Improve Psychotherapy Research and Practice? Adm Policy Ment Health. 2020;47(5):852-5.

22. Rocheteau E. On the role of artificial intelligence in psychiatry. Br J Psychiatry. 2023;222(2):54-7.

23. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85-117.

24. Yang GR, Wang XJ. Artificial Neural Networks for Neuroscientists: A Primer. Neuron. 2020;107(6):1048-70.

25. McGowan J, Straus S, Moher D, Langlois EV, O'Brien KK, Horsley T, et al. Reporting scoping reviews—PRISMA ScR extension. Journal of clinical epidemiology. 2020;123:177-9.

26. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.

27. Gori A, Lauro-Grotto R, Giannini M, Schuldberg D. Predicting treatment outcome by combining different assessment tools: Toward an integrative model of decision support in psychotherapy. Journal of Psychotherapy Integration. 2010;20(2):251-69.

28. Nitti M, Ciavolino E, Salvatore S, Gennaro A. Analyzing psychotherapy process as intersubjective sensemaking: An approach based on discourse analysis and neural networks. Psychotherapy Research. 2010;20(5):546-63.

29. Koppe G, Guloksuz S, Reininghaus U, Durstewitz D. Recurrent Neural Networks in Mobile Sampling and Intervention. Schizophr Bull. 2019;45(2):272-6.

30. Ewbank MP, Cummins R, Tablan V, Bateup S, Catarino A, Martin AJ, et al. Quantifying the Association Between Psycho-therapy Content and Clinical Outcomes Using Deep Learning. JAMA Psychiatry. 2020;77(1):35-43.

31. Burger F, Neerincx MA, Brinkman WP. Natural language processing for cognitive therapy: Extracting schemas from thought records. PLoS One. 2021;16(10):e0257832.

32. Bennemann B, Schwartz B, Giesemann J, Lutz W. Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. The British Journal of Psychiatry. 2022;220(4):192-201.

33. Chen M, Liang X, Xu Y. Construction and Analysis of Emotion Recognition and Psychotherapy System of College Students under Convolutional Neural Network and Interactive Technology. Comput Intell Neurosci. 2022;2022:5993839.

34. Rodrigo H, Beukes EW, Andersson G, Manchaiah V. Predicting the Outcomes of Internet-Based Cognitive Behavioral Therapy for Tinnitus: Applications of Artificial Neural Network and Support Vector Machine. Am J Audiol. 2022;31(4):1167-77.

35. Drayton M. The Minnesota Multiphasic Personality Inventory-2 (MMPI-2). Occup Med (Lond). 2009;59(2):135-6.

36. Basaraba CN, Scodes JM, Dambreville R, Radigan M, Dachepally P, Gu G, et al. Prediction Tool for Individual Outcome Trajectories Across the Next Year in First-Episode Psychosis in Coordinated Specialty Care. JAMA Psychiatry. 2023;80(1):49-56.

37. Coley RY, Jennifer MB, Arne B, Gregory ES. Predicting outcomes of psychotherapy for depression with electronic health record data. Journal of Affective Disorders Reports. 2021;6:100198.

38. Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry. 2021;20(2):154-70.

39. Lutz W, Zimmermann D, Müller VNLS, Deisenhofer A-K, Rubel JA. Randomized controlled trial to evaluate the effects of personalized prediction and adaptation tools on treatment

outcome in outpatient psychotherapy: study protocol. BMC Psychiatry. 2017;17(1):306.

40. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. Journal of the American Medical Informatics Association. 2019;26(6):561-76.

41. Rezaeenour J, Ahmadi M, Jelodar H, Shahrooei R. Systematic review of content analysis algorithms based on deep neural networks. Multimedia Tools and Applications. 2023;82(12):17879-903.

42. Gottschalk LA. The application of computerized content analysis of natural language in psychotherapy research now and in the future. Am J Psychother. 2000;54(3):305-11.

43. Lu X. Deep Learning Based Emotion Recognition and Visualization of Figural Representation. Front Psychol. 2021;12:818833.

44. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms. JMIR Ment Health. 2021;8(10):e22651.

45. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach. Health Informatics Journal. 2022;28(4):14604582221142442.

46. Hossain E, Rana R, Higgins N, Soar J, Barua PD, Pisani AR, et al. Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. Comput Biol Med. 2023;155:106649.

Figures and Tables

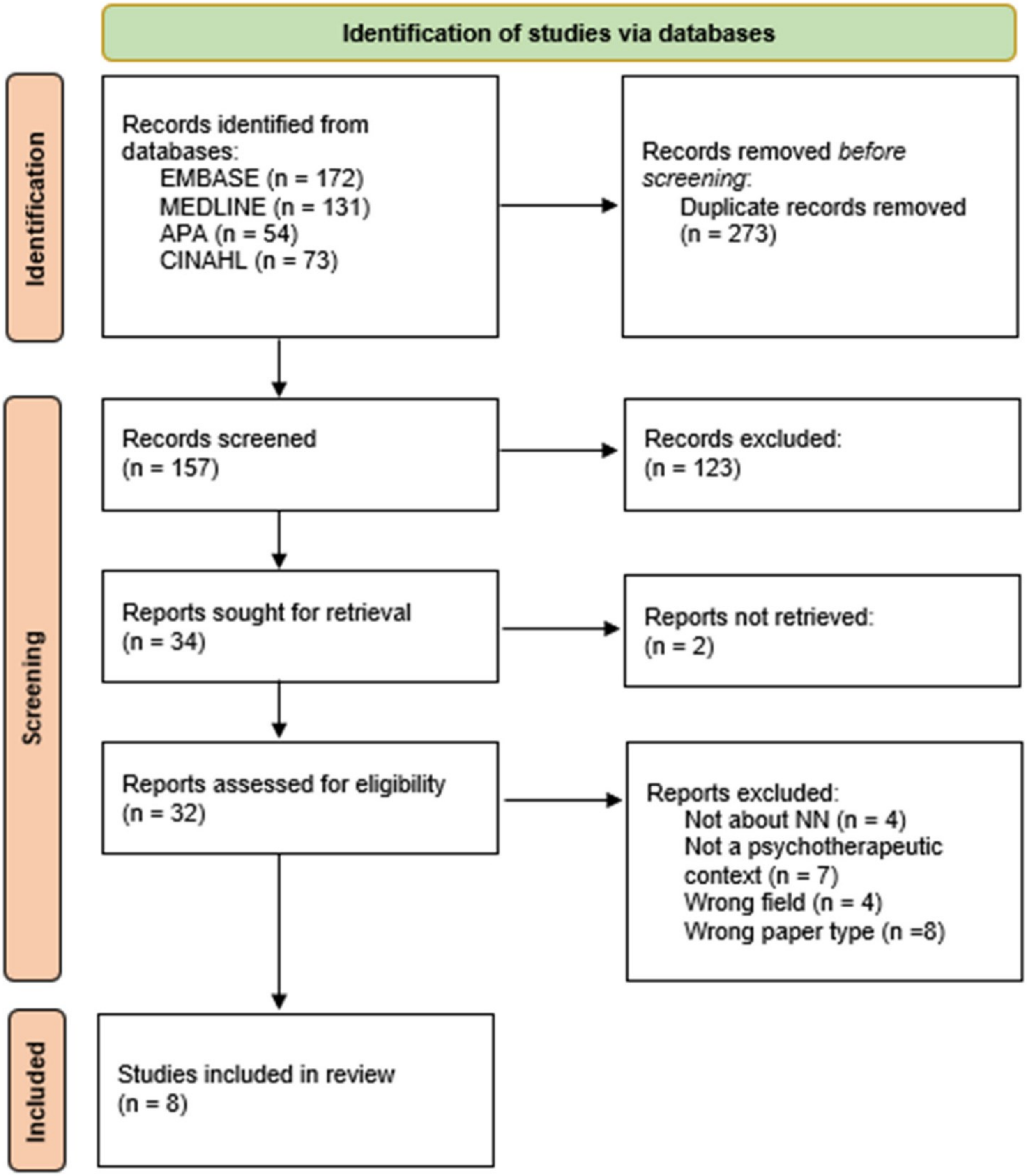
Identification of studies via databases
Identification
Screening
Included
Records identified from databases:
EMBASE (n = 172)
MEDLINE (n = 131)
APA (n = 54)
CINAHL (n = 73)
Records removed *before screening*:
Duplicate records removed (n = 273)
Records screened (n = 157)
Records excluded: (n = 123)
Reports sought for retrieval (n = 34)
Reports not retrieved: (n = 2)
Reports assessed for eligibility (n = 32)
Reports excluded:
Not about NN (n = 4)
Not a psychotherapeutic context (n = 7)
Wrong field (n = 4)
Wrong paper type (n =8)
Studies included in review (n = 8)

Figure 1. PRISMA flowchart for the inclusion of studies.

Table 1. Scoping review study selection detailed results

Article	**Population**	**Neural Network**	**Intervention**	**Metrics used**	**Main outcome**	**Quality assessment**
(Gori et al., 2010) (27)	Patients who requested psychotherapeutic treatment (n = 150)	Artificial Neural Networks	Predict treatment outcome	Italian version of the MMPI-2	ANN forecasted 81% of the clinical outcomes (successful/unsuccessful therapy)	Low
(Nitti et al., 2010) (28)	Patient who followed a psychodynamic psychotherapy (n=1)	Discourse Flow Analysis	Analysis of the verbatim of 43 sessions of therapy	Indexes of discourse network (connectivity, activity, and regulation)	Neural networks allow us the identification of patterns characterizing the psychotherapy process.	Very low
(Koppe et al., 2019)(29)	Experiences and context specific interventions for psychosis. (n =n/a)	Recurrent Neural Networks	Mobile sampling for prediction of symptoms	None	RNNs could be used to forecast individual trajectories and schedule online feedback and interventions.	Very low
(Ewbank et al., 2020)(30)	Patients who followed internet-enabled CBT (n=13 073)	Bidirectional LSTM	Automated categorization of therapist utterances	PHQ-9, GAD-7	The model achieved acceptable categorization and has reached human-level agreement.	Moderate
(Burger et al., 2021)(31)	Healthy participants (n=320)	Recurrent Neural Networks	Identifying schemas (derived from schema therapy)	HDAS, BDI-IA, CDS	Schemas can be automatically extracted and NN perform betweer than KNN and support vector	Low

			from thought records.		approaches.	
(Bennemann et al., 2022)(32)	Outpatients treated with CBT (n=2543)	Ensemble modeling using several machine learning algorithms (including artificial neural networks)	Predicting the drop-out rates of patients	PSSI, BSI	Neural networks were identified to be less suited to predict naturalistic data-sets and binary events.	Moderate
(Chen et al., 2022)(33)	Students who followed human-computer interaction psychotherapy (n = 120)	Convolational Neural Networks	Recognizing emotions based on the human-computer interaction	Kaggle facial emotion recognition dataset	Convulational neural networks are better to recognize student emotions than backpropagation neural network and decision tree algorithms.	Low
(Rodrigo et al., 2022)(34)	Individuals who completed internet-enabled CBT for tinnitus (n=228)	Artificial Neural Networks	Predicting the treatment outcome by determining variables associated to treatment success.	TFI	The best predictive model was achieved by the artificial neural network with an area under the curve with a value of 0.73 over 33 predictor variables.	Low

Table A1. Electronic search planification strategy for the systematic review conducted.

	Neural Networks	**Psychotherapy**
EMBASE	X	psychotherapy/ OR psychology
MEDLINE	X	Psychology Services, Psychology/ or Psychotherapy, Psychiatric
APA	Neural networks/artificial intelligence	psychotherapy/ or psychiatry/ or mental health/
CINHAL	X	
Free vocabulary	(((Neural networks OR machine learning OR deep learning OR artificial intelligence) N3 (Neural OR neuron OR networks OR computer science OR deep OR natural language processing))	(psychotherapy OR psychology) N2 (interventions* OR treatment* OR psychoed* OR psychology* OR psychiatry* OR psych*)

Article 4. Comparing the Performance of Machine Learning Algorithms in the Automatic Classification of Psychotherapeutic Interactions in Avatar Therapy

Alexandre Hudon

Kingsada Phraxayavong

Stéphane Potvin

Alexandre Dumais

Publié dans *Machine Learning & Knowledge Extraction (2023).*
Article révisé par les pairs.

Abstract

(1) Background: Avatar Therapy (AT) is currently being studied to help patients suffering from treatment-resistant schizophrenia. Facilitating annotations of immersive verbatims in AT by using classification algorithms could be an interesting avenue to reduce the time and cost of conducting such analysis and adding objective quantitative data in the classification of the different interactions taking place during the therapy. The aim of this study is to compare the performance of machine learning algorithms in the automatic annotation of immersive session verbatims of AT. (2) Methods: Five machine learning algorithms were implemented over a dataset as per the Scikit-Learn library: Support vector classifier, Linear support vector classifier, Multinomial Naïve Bayes, Decision Tree, and Multi-layer perceptron classifier. The dataset consisted of the 27 different types of interactions taking place in AT for the Avatar and the patient for 35 patients who underwent eight immersive sessions as part of their treatment in AT. (3) Results: The Linear SVC performed best over the dataset as compared with the other algorithms with the highest accuracy score, recall score, and F1-Score. The regular SVC performed best for precision. (4) Conclusions: This study presented an objective method for classifying textual interactions based on immersive session verbatims and gave a first comparison of multiple machine learning algorithms on AT.

Keywords

artificial intelligence; virtual reality therapy; auditory hallucinations; schizophrenia; psychotherapy; machine learning; algorithms

Introduction

A severe mental disorder such as schizophrenia has a high social burden [1]. The economic burden of schizophrenia in the United States alone reached 155.7 billion dollars in 2013 [2]. The mental state of those suffering from schizophrenia may be disturbed. This disturbance can include delusions and hallucinations, also known as positive symptoms. Patients with schizophrenia are more likely to experience auditory hallucinations [3]. A thorough strategy is therefore required for the treatment of positive symptoms. Psychoeducation is used to explain the diagnosis, and psychopharmacological treatments are added to deal with delusions and hallucinations [4,5]. Despite receiving regular medical treatments, over 25% of individuals still have positive symptoms [6,7]. Antipsychotic drugs and psychotherapy techniques such as family interventions, psychoeducation, and cognitive-behavioral therapy (CBT) are frequently used in the standard of care treatment [8,9].

Novel therapies such as Avatar Therapy (AT) emerged to account for this problem and offer an alternate solution for patients suffering from schizophrenia with refractory auditory hallucinations [10]. This therapy is still being studied to validate its efficiency in reducing patients' refractory auditory hallucinations and assessing their wellbeing. Avatar Therapy implies the use of a virtual reality headset where the therapists interact with the patient in an immersive environment [11]. In the environment, the therapist animates a visual representation (pre-configured by the patient) of the patient's auditory hallucination. AT was initially developed by Leff et al. (2014) in 2008 [12]. In their first pilot trial for this type of therapy, AT consisted of 7 weeks of therapy (one session per week), comprising six immersive 30 min sessions with the Avatar. This trial enrolled 26 patients, 16 received AT, and they benefited from a significant reduction in the frequency and intensity of their auditory hallucinations [13]. Furthermore, it highlighted a significant reduction in depressive symptoms. In 2016, Craig and al. (2018, trial number: ISRCTN, number 65,314 790) conducted the first single-blind, randomized controlled trial with 150 patients from 18 to 65 years who had received a clinical diagnosis of schizophrenia spectrum and had auditory verbal hallucinations despite continued treatment [14]. These patients were randomly assigned to receive AT or supportive therapy. The main outcome was reduction in auditory verbal hallucinations at 12 weeks on the Psychotic Symptoms Rating Scales Auditory Hallucinations (PSYRATS-AH) [14]. At the Institut Universitaire en Santé Mentale de l'Université de Montréal (IUSMM), an undergoing clinical trial piloted by Dr. Dumais and Dr. Potvin is comparing AT to CBT for patients suffering from schizophrenia with auditory hallucinations under continued treatment. The trial includes 136 participants: 68 undergoing AT and 68 undergoing CBT. While this trial is underway, a one-year pilot randomized comparative trial evaluating the short- and long-term efficacity of VRT over CBT at the IUSMM for this population and assessed 37 patients who undertook AT and 37 who undertook CBT [15]. AT achieved larger effect sizes than CBT on auditory hallucinations for these patients as well as showed significant results on persecutory beliefs and quality of life [15].

While clinical trials are showing promising outcomes regarding the impact of Avatar Therapy (AT) in reducing auditory hallucinations among individuals with schizophrenia, a few studies have attempted to qualitatively assess the verbatims of immersive sessions to gain a deeper understanding of the therapeutic process. Commonly employed techniques for this assessment include content analysis of therapeutic sessions, semi-structured interviews, and questionnaires. However, these methods can be time-consuming, require significant human resources, are susceptible to biases depending on the analytical approach taken, and may be hard to generalize [16]. These biases include misclassification of outcomes, selection biases, and confounding biases [17]. Often, they focus on a limited set of items, which makes it challenging to obtain a comprehensive understanding of the underlying therapeutic process. Qualitative approaches such as phenomenology or grounded theory are often utilized to explore the nuances of therapeutic sessions [18].

In 2018, an initial content analysis of AT was conducted, examining the therapeutic sessions of 12 patients who underwent the therapy [19]. They analyzed up to 84 immersive session verbatims until reaching a saturation point. This analysis revealed five thematic areas that emerged from patients' dialogue with the Avatar: emotional response to voices, beliefs about voices and schizophrenia, self-perceptions, coping mechanisms, and aspirations [19]. These themes provided initial insights into potential therapeutic targets in AT. Building upon this, Beaudoin et al. conducted a subsequent study in 2021, qualitatively assessing 125 therapy verbatims (totaling 1419 min) from 18 patients [20]. The aim was to gain a deeper understanding of the dynamics between the patient and the Avatar. Two major key themes were identified for the Avatar: confrontational techniques (comprising eight sub-themes) and positive techniques (comprising six sub-themes). For the patients, five key themes were identified: self-perceptions, emotional responses, aspirations, coping mechanisms, and beliefs about voices and schizophrenia. These five themes encompassed a total of 14 sub-themes [20]. These qualitative studies contribute to the knowledge of the therapeutic process in AT, shedding light on the interactions between patients and Avatars and identifying key thematic areas that could guide future research and therapeutic interventions. While qualitative data can be informative and extensive in nature, it lacks the quantitative counterpart necessary to determine the specific elements of therapy that may contribute to positive outcomes.

Classification algorithms are often used in the field of medicine to account for this lack of quantitative assessment [21]. As an example, a study designed by Chekroud et al. reviewed the use of classification algorithms to predict treatment outcomes in psychiatry, ranging from medication to psychotherapies to digital interventions and neurobiological treatments, and included the classification of text entities [22]. They conclude that the use of classification algorithms is a new but important approach to improving the effectiveness of mental health care [22]. In mental health, few of these approaches have been attempted, mostly due to the limited amount of data available (e.g., a small number of therapeutic verbatims). In Avatar Therapy, the complexity of having interactions between three individuals and the fact that it is less readily available to the public limits the extent of usable data for constructing a database. As an example, this can yield databases that are smaller than data readily available for internet-based CBT. A classification algorithm applicable to small databases is therefore needed for such cases. A recent review assessed machine learning algorithms used in the context of psychiatry, psychology, and social sciences and identified several potential algorithms that can be used with small datasets [23]. Classification algorithms such as Naïve Bayes, Decision Tree, and support vector machine classifiers were found to be relevant in these contexts. According to the identified algorithms, the most used and best-performing algorithm is the support vector machine [23]. This opens the door to merging previous content analysis with quantifiable data to forecast the prediction of therapeutic outcomes in the context of psychotherapy. Facilitating annotations of immersive verbatims in AT by using classification algorithms could be an interesting avenue to reduce the time and cost of conducting such analysis and adding objective quantitative data in the identification and classification of the different interactions taking place during the therapy.

The aim of this study is to compare the performance of machine learning algorithms in the automatic annotation of immersive session verbatims of AT. Considering the resources required to conduct such a task and the subjectivity of manual annotation of psychotherapy verbatims, the use of AI algorithms may be an interesting avenue. The main goal to be achieved in this study is to be able to identify the best-performing algorithm to conduct automated annotations of AT verbatims. This requires the proper identification of the best-performing algorithm for the specific context of AT. We hypothesize that support vector machine algorithms will perform best considering the limited dataset available for AT at this time and considering the high number of features being integrated for the automated classification of the interactions taking place in the verbatims.

Materials and Methods

Participants and Recruitment

The data utilized in this study originated from individuals who participated in pilot trials conducted at the Centre de recherche de l'Institut universitaire en santé mentale de Montréal (CR-IUSMM) and an ongoing trial that compares AT to CBT. These participants were enrolled in the clinical trial registered on Clinicaltrials.gov, identified by the number NCT03585127 [15]. All participants received a total of nine one-hour psychotherapeutic sessions, of which eight were immersive sessions involving interaction with a virtual representation of their auditory verbal hallucinations—the Avatar. The participants included in this study were patients of the IUSMM aged over 18 years. They all suffered from treatment-resistant schizophrenia (TRS), defined by the lack of response to two or more dopaminergic antagonists as expressed by the persistence of auditory hallucinations. The AT sessions were administered between the years 2017 and 2022.

Dataset: Corpus of Avatar Therapy and Features

Immersive sessions of 35 patients who had undergone AT were transcribed verbatim from audio recordings by research auxiliaries. The verbatims were then verified by AH to ensure the integrity of the transcriptions. This yielded 288 verbatims representing over 250 h of immersion in AT. Annotations of the interactions between the patients and the Avatars were classified as per the 27 themes described in Beaudoin et al. 2021 [20]. The themes are presented in Table 1 for the Avatar and Table 2 for the patients.

-- Please insert Table 1 here --

-- Please insert Table 2 here --

A dataset comprising 280 therapy transcripts from thirty-five randomly selected patients who underwent Avatar Therapy (AT) between 2017 and 2022 at our institution was compiled. Each patient participated in eight therapy sessions, resulting in an average of eight transcripts per patient. The transcripts were originally manually typed and were in Canadian French. For annotation purposes, the transcripts were manually annotated using the 27 themes described in the study conducted by Beaudoin et al. in 2021 [20]. The annotation process was carried out using QDA Miner version 5, a qualitative data analysis software developed by Provalis Research [24]. The annotations were subsequently extracted as text files, with each file containing a varying number of interactions (ranging from 1 to 40) related to the same theme. These extracted annotations were then categorized into two conceptual databases: Avatar and Patient, following the representation depicted in Figure 1.

-- Please insert Figure 1 here --

Machine learning algorithms

Five algorithms for automated text classification were implemented over the AT dataset in Python 3.11 as per the classification identified in the previous literature review for the context of psychotherapy: Support vector classifier (SVC), Linear support vector classifier (Linear SVC), Multinomial Naïve Bayes (Multinomial NB), Decision Tree (DT), and Multi-layer perceptron classifier (MLP) [23]. They were all used over the Avatar conceptual dataset and the Patient conceptual dataset. A GridSearchCV (GSCV) technique from the Scikit-Learn library was employed to optimize the performance of the machine learning algorithm and improve classification strategies. GSCV is a valuable tool as it allows users to explore various hyperparameters and cross-validate the classifier's predictions, thereby identifying the optimal combination of parameters that yield the best performance. In this study, GSCV was applied to both SVC and LSVC classifiers [25]. Default parameters were utilized for the DT, MLP, and Multinomial NB classifiers, as they demonstrated superior performance when considering hyperparameterization.

The algorithms were paired with a term frequency-inverse document frequency (TF-IDF) statistic, known for its superior performance in text classification when compared with other algorithm-tokenizer combinations. To implement TF-IDF tokenization, we selected the TfidfVectorizer provided by the Scikit-Learn library. This vectorizer facilitates the conversion of the raw text extracted from the interview's interactions into numerical vectors [26]. Additionally, vectorizers can be customized to accommodate stop-words if necessary. Because the classification categories were designed to separate text entities based on their distinct intrinsic characteristics the assumption is that the features are linearly separable [20].

Support Vector Classifier (SVC)

A Support vector classifier is employed for supervised classification tasks [27]. Finding the best hyperplane to divide several classes of data points in a high-dimensional feature space is the main goal of this particular support vector machine (SVM) approach [28]. Maximizing the margin between classes, it does this with the intention of achieving good generalization performance [29]. It operates by locating a subset of training samples known as support vectors that serve as the decision boundary's key points. These support vectors are critical in choosing the best hyperplane because they are located closest to the decision boundary.

The implementation used for the SVC in this study is from Scikit-Learn, more precisely, the SVC class of the SVM library [26,30].

Linear Support Vector Classifier (Linear SVC)

The Linear support vector classifier belongs to the family of support vector machines. As compared with SVC, Linear SVC uses a linear kernel. A kernel is a mathematical function that is used in a variety of machine-learning methods to turn data into a higher-dimensional feature space [31]. The ability of algorithms to address complicated issues that can be challenging or even impossible to handle in the original input space is fundamentally dependent on kernels. Therefore, a linear kernel is used when the data are linearly separable.

The implementation used for the SVC in this study is from Scikit-Learn, more precisely, the SVC class of the SVM library with the specification of using a linear kernel [30,32].

Multinomial Naïve Bayes Classifier (Multinomial NB)

The main application of the probabilistic machine learning technique known as the Multinomial Naïve Bayes classifier is text classification problems. It is a development of the Naïve Bayes method, which relies on the Bayes theorem and assumes that the characteristics are conditionally independent of the class [33]. The Bayes theorem enables us to revise the likelihood that Event A will occur considering novel data or supporting evidence provided by Event B. By combining the prior probability (P(A)) and the likelihood (P(B|A)), it offers a method for calculating the posterior probability (P(A|B)) [34]. To handle discrete features in text data, such as word counts or frequencies, the Multinomial Naïve Bayes classifier was developed.

The implementation used for the SVC in this study is from Scikit-Learn, more precisely, the MultinomialNB class of the Naïve Bayes library [30].

Decision Tree Classifier (DT)

Decision Tree-based classifiers are non-parametrized and utilized as supervised learning methods for item classification. These classifiers represent observations about an item through branches and draw conclusions about the item's value or score through leaves [35]. The splitting of observations across branches is determined by predefined rules based on the categories used for classification. In the context of text classification, the underlying concept is that each piece of text being classified undergoes a process of splitting across branches until it reaches a leaf (representing a category) according to probabilistic rules established by the designer of the Decision Tree [36].

The implementation used for the DT in this study is from Scikit-Learn, more precisely, the DecisionTreeClassifier class [30].

Multi-Layer Perceptron Classifier (MLP)

A Multi-layer perceptron classifier is used for a variety of machine learning tasks, including classification. It is a model of a feedforward neural network made up of numerous layers of coupled neurons [37]. The input layer, one or more hidden layers, and the output layer are commonly present in the layered structure of the MLP classifier. Multiple neurons make up each layer, which executes calculations on the incoming data and relays the results to the following layer. Each neuron in each layer of an MLP is connected to every other neuron in the neighboring layers, indicating that the MLP is fully connected. Weights attached to the connections between neurons govern the strength and significance of the information moving through the network [38].

The implementation used for the MLP in this study is from Scikit-Learn, more precisely, the MLPClassifier class from the neural_network library [30].

Data analysis and validation

A partitioning strategy was employed for each conceptual database, where 70% of the annotated documents were used for training the algorithms, while the remaining 30% were utilized for testing purposes [39]. The objective was to establish a statistical probability for each algorithm, represented by a predictive score, indicating the adequacy of classifying an interaction. The training and testing sets were intentionally non-overlapping to adhere to recommended design practices [40,41]. The predictive score corresponds to the average accuracy, measured by the F1-Score, of the themes being evaluated during testing. Additionally, a tenfold cross-validation technique was implemented using the K-Fold model from the Scikit-Learn library for each algorithm [30,42].

The Classification Report tool from the Scikit-Learn metrics module was utilized to gather information regarding the classification performance of each theme, including the precision, recall, and F1-Score for each algorithm. Precision represents the positive predictive value, recall indicates the sensitivity of the prediction, and the F1-Score reflects the accuracy of theme classification [43]. The F1-Score is a commonly used measure in text classification that strikes a balance between precision and recall, providing an overall assessment of classification accuracy. The F1-Score is, therefore, the harmonic mean between precision and recall [44].

Results

Sample characteristics

Interactions taking place in the verbatims of 35 patients were used by the five machine learning algorithms in this study to conduct automated annotation. The characteristics of the sampled patients are found in Table 3.

-- Please insert Table 3 here --

Performance of Machine Learning Algorithms

The average performance of the machine learning algorithm for the automatic annotation of the verbatim is found in Table 4. It can be observed that the Linear SVC performs best over the dataset as compared with the other algorithms with the highest accuracy score, recall score, and F1-Score. The regular SVC performs best for precision over the dataset. Overall, the DT classifier performs the worst over the analyzed metrics. Descriptive visualization of the F1-Score comparisons can be observed in Figure 2.

-- Please insert Figure 2 here --
-- Please insert Table 4 here --

The average performances of the different classifiers are presented in Table 5. As for the performance on the Avatar database, it can be observed that the Linear SVC performs best for the F1-Score as well as all the other metrics except for the precision, where the regular SVC offers superior performance. The Decision Tree performs poorly over the database with the smallest F1-Score. Descriptive visualization of the F1-Score comparisons of the models over the Patient dataset can be observed in Figure 3.

-- Please insert Table 5 here --
-- Please insert Figure 3 here--

Discussion

This study aimed to compare the performance of machine learning algorithms in the automatic annotation of immersive session verbatims of AT. From the five implementations of machine algorithms over both the Avatar and Patient conceptual databases, it was observed that the Linear SVC performed the best across all metrics except for the precision. The regular SVC performed best for the precision metrics.

Artificial intelligence, especially the field of machine learning, could therefore provide an interesting avenue for automated annotations of psychotherapeutic verbatims, which are usually performed by human coders. This would have the potential to save resources (cost and time) as well as balance subjectivity biases introduced by qualitative assessment of verbatims. Such techniques should be further explored.

While few implementations of supervised machine learning algorithms exist in the clinical applications of psychiatry and psychotherapy, text classification and automated annotation is used in different aspects of medicine. A study by Gibbons et al. (2017) tackled the challenge of classifying open-text feedback of doctor performances with human-level accuracy on a corpus of 1636 open-text comments relating to the performance of 548 doctors [45]. With a dataset of comparable size as the one used in our study, it was found that their support vector machine classifier (SVM) had a similar F1-Score performance as the one observed in AT. However, in their implementation, DT and the combinations of three and more models yielded better overall performance. This can be explained by the context of their applications of machine learning algorithms' performance comparison, considering they used a context of an open-ended survey as their corpus, which comprised fewer features than the ones used in AT. As complexity grows, algorithms such as SVM-based classifiers perform better in the context of textual entities classified over more features [46,47].

The performance of LSVC over SVC in the context of AT might be intrinsic to the linear separation of the different themes [48]. Considering the previous qualitative analysis conducted on AT, the themes identified were attempted to be as linearly separable as possible. This can explain the overall poor performance of DT and Multinomial NB. A recent review of the application of machine learning algorithms on text classification highlights that Naïve Bayes algorithms often perform poorly, as they assume that all the features are entirely independent of each other, which often is not the case when the corpus is human-generated such as in the context of AT [49]. The Multinomial NB assumes a multinomial distribution of AT interactions that might not be accurate [50]. As for DT, continuous data such as the dataset of this study offers many branching, and this can lead to poor performances. As for the precision performance of SVC over Linear SVC, SVC with an appropriate non-linear kernel can provide better precision by capturing the underlying complexities of the data. The data in AT refers to interactions between the Patient and the Avatar and is intrinsically complex as defined by the underlying naturalistic language being assessed.
Finally, the performance of the MLP might have been impacted by the small size of the database. Neural network algorithm often needs a vast array of data to achieve adequate performance [51].

Limitations

The current analysis of the performance for the different implementations of the machine learning algorithms as described is limited by the small database offered by AT. As more patients are included in the dataset, the trend of the performances for the different algorithms will be re-assessed. It is also important to mention that the transcripts examined in this study were written in Canadian French. A challenge was encountered in finding vectorizers that incorporated stop-words specifically for the Canadian French language. Stop-words are words that are typically excluded from the tokenization process as they hold little or no significant meaning. The absence of appropriate stop-words for Canadian French can potentially impact the accuracy of the analysis, as it may result in insignificant words being included in the word vectors and affecting the overall results.

Conclusions

To conclude, this study compared the performances of five machine learning algorithms over the AT dataset. More precisely, it focused on the classification of textual interactions from verbatims of patients suffering from TRS undergoing immersive virtual reality sessions in AT. The Linear SVC algorithm was identified as being the algorithm that performed best in terms of the accuracy, recall, and F1-Score for the Avatar conceptual dataset and the Patient conceptual dataset. The SVC algorithm also performed well compared with the other algorithm, achieving the best performances for precision. This study offers a first comparison of several machine learning algorithms on AT and provides an objective approach to the classification of textual interactions based on immersive session verbatims. Future studies could use this approach to provide insight relating to the elements being classified and the therapeutical response of patients as per their experience with AT immersive sessions.

Author Contributions

Conceptualization, A.H., K.P., S.P. and A.D.; methodology, A.H. and A.D.; validation, A.H. and A.D.; formal analysis, A.H.; investigation, A.H.; data curation, A.H.; writing—original draft preparation, A.H.; writing—review and editing, A.H., K.P., S.P. and A.D.; supervision, K.P., S.P. and A.D.; project administration, K.P.; funding acquisition, K.P., S.P. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work was indirectly supported by Le Fonds de recherche du Québec—Santé (FRQS); Otsuka Canada Pharmaceutical Inc.; Chaire Eli Lilly Canada de recherche en schizophrénie; MEI (Ministère de l'Économie et de l'Innovation); Services et recherches psychiatriques AD; and Fonds d'excellence en recherche Apogée Canada. These funding bodies had no part in the data collection, analysis, interpretation of data, or in writing the manuscript.

Institutional Review Board Statement

This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients. Patients that are part of this study were selected based on the proof-of-concept trial from Percy du Sert's 2018 study and Dellazizzo's 2021 study [15]. The trial was conducted in accordance with the Declaration of Helsinki and was approved by the institutional ethical committee (CER IPPM 16-17-06). We obtained written informed consent from all patients.

Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available due to patients' privacy but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Charlson, F.J., et al., Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. Schizophr Bull, 2018. 44(6): p. 1195-1203.

2. Cloutier, M., et al., The Economic Burden of Schizophrenia in the United States in 2013. J Clin Psychiatry, 2016. 77(6): p. 764-71.

3. Habtewold, T.D., et al., Deep Clinical Phenotyping of Schizophrenia Spectrum Disorders Using Data-Driven Methods: Marching towards Precision Psychiatry. Journal of Personalized Medicine, 2023. 13(6): p. 954.

4. Huhn, M., et al., Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. Lancet, 2019. 394(10202): p. 939-951.

5. Xia, J., L.B. Merinder, and M.R. Belgamwar, Psychoeducation for schizophrenia. Cochrane Database Syst Rev, 2011. 2011(6): p. Cd002831.

6. Lally, J., et al., Treatment-resistant schizophrenia: current insights on the pharmacogenomics of antipsychotics. Pharmgenomics Pers Med, 2016. 9: p. 117-129.

7. Potkin, S.G., et al., The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. npj Schizophrenia, 2020. 6(1): p. 1.

8. Stępnicki, P., M. Kondej, and A.A. Kaczor, Current Concepts and Treatments of Schizophrenia. Molecules, 2018. 23(8): p. 2087.

9. Guaiana, G., et al., Cognitive behavioural therapy (group) for schizophrenia. Cochrane Database Syst Rev, 2022. 7(7): p. Cd009608.

10. Aali, G., T. Kariotis, and F. Shokraneh, Avatar Therapy for people with schizophrenia or related disorders. Cochrane Database Syst Rev, 2020. 5(5): p. Cd011898.

11. Dellazizzo, L., et al., Avatar Therapy for Persistent Auditory Verbal Hallucinations in an Ultra-Resistant Schizophrenia Patient: A Case Report. Front Psychiatry, 2018. 9: p. 131.

12. Leff, J., et al., Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? Psychosis, 2014. 6(2): p. 166-176.

13. Leff, J., et al., Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. Br J Psychiatry, 2013. 202: p. 428-33.

14. Craig, T.K., et al., AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry, 2018. 5(1): p. 31-40.

15. Dellazizzo, L., et al., One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. NPJ Schizophr, 2021. 7(1): p. 9.

16. Chai, H.H., et al., A Concise Review on Qualitative Research in Dentistry. Int J Environ Res Public Health, 2021. 18(3).

17. Pannucci, C.J. and E.G. Wilkins, Identifying and avoiding bias in research. Plast Reconstr Surg, 2010. 126(2): p. 619-625.

18. Starks, H. and S.B. Trinidad, Choose your method: a comparison of phenomenology, discourse analysis, and grounded theory. Qual Health Res, 2007. 17(10): p. 1372-80.

19. Dellazizzo, L., et al., Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychother, 2018. 25(6): p. 878-885.

20. Beaudoin, M., et al., The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clin Psychol Psychother, 2021. 28(3): p. 500-518.

21. Sidey-Gibbons, J.A.M. and C.J. Sidey-Gibbons, Machine learning in medicine: a practical introduction. BMC Medical Research Methodology, 2019. 19(1): p. 64.

22. Chekroud, A.M., et al., The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry, 2021. 20(2): p. 154-170.

23. Hudon, A., et al., Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms. JMIR Ment Health, 2021. 8(10): p. e22651.

24. Lewis, R.B. and S.M. Maas, QDA Miner 2.0: Mixed-model qualitative data analysis software. Field methods, 2007. 19(1): p. 87-108.

25. Paper, D. and D. Paper, Scikit-Learn Classifier Tuning from Simple Training Sets. Hands-on

Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python, 2020: p. 137-163.

26. Komer, B., J. Bergstra, and C. Eliasmith, Hyperopt-sklearn. Automated Machine Learning: Methods, Systems, Challenges, 2019: p. 97-111.

27. Mammone, A., M. Turchi, and N. Cristianini, Support vector machines. Wiley Interdisciplinary Reviews: Computational Statistics, 2009. 1(3): p. 283-289.

28. Shao, Y.-H., W.-J. Chen, and N.-Y. Deng, Nonparallel hyperplane support vector machine for binary classification problems. Information Sciences, 2014. 263: p. 22-35.

29. Xu, J., et al. Multi-class support vector machine via maximizing multi-class margins. in The 26th International Joint Conference on Artificial Intelligence (IJCAI 2017). 2017.

30. Pedregosa, F., et al., Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 2011. 12: p. 2825-2830.

31. Almaiah, M.A., et al., Performance investigation of principal component analysis for intrusion detection system using different support vector machine kernels. Electronics, 2022. 11(21): p. 3571.

32. Varoquaux, G., et al., Scikit-learn: Machine learning without learning the machinery. GetMobile: Mobile Computing and Communications, 2015. 19(1): p. 29-33.

33. Rish, I. An empirical study of the naive Bayes classifier. in IJCAI 2001 workshop on empirical methods in artificial intelligence. 2001.

34. Berrar, D., Bayes' theorem and naive Bayes classifier. Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics, 2018. 403: p. 412.

35. Kingsford, C. and S.L. Salzberg, What are decision trees? Nature biotechnology, 2008. 26(9): p. 1011-1013.

36. Kotsiantis, S.B., Decision trees: a recent overview. Artificial Intelligence Review, 2013. 39: p. 261-283.

37. Ramchoun, H., et al., Multilayer perceptron: Architecture optimization and training. 2016.

38. Popescu, M.-C., et al., Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems, 2009. 8(7): p. 579-588.

39. Gholamy, A., V. Kreinovich, and O. Kosheleva, Why 70/30 or 80/20 relation between

training and testing sets: A pedagogical explanation. 2018.

40. Bhavsar, H. and A. Ganatra, A comparative study of training algorithms for supervised machine learning. International Journal of Soft Computing and Engineering (IJSCE), 2012. 2(4): p. 2231-2307.

41. Huang, X., G. Jin, and W. Ruan, Machine Learning Basics, in Machine Learning Safety. 2012, Springer. p. 3-13.

42. Wong, T.-T. and P.-Y. Yeh, Reliable accuracy estimates from k-fold cross validation. IEEE Transactions on Knowledge and Data Engineering, 2019. 32(8): p. 1586-1594.

43. Goutte, C. and E. Gaussier. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. in European conference on information retrieval. 2005. Springer.

44. Opitz, J. and S. Burst, Macro f1 and macro f1. arXiv preprint arXiv:1911.03347, 2019.

45. Gibbons, C., et al., Supervised Machine Learning Algorithms Can Classify Open-Text Feedback of Doctor Performance With Human-Level Accuracy. J Med Internet Res, 2017. 19(3): p. e65.

46. Joachims, T. Text categorization with support vector machines: Learning with many relevant features. in European conference on machine learning. 1998. Springer.

47. Liu, Z., et al. Study on SVM compared with the other text classification methods. in 2010 Second international workshop on education technology and computer science. 2010. IEEE.

48. Amarappa, S. and S. Sathyanarayana, Data classification using Support vector Machine (SVM), a simplified approach. Int. J. Electron. Comput. Sci. Eng, 2014. 3: p. 435-445.

49. Li, R., et al. A Review of Machine Learning Algorithms for Text Classification. 2022. Singapore: Springer Nature Singapore.

50. Harzevili, N.S. and S.H. Alizadeh, Mixture of latent multinomial naive Bayes classifier. Applied Soft Computing, 2018. 69: p. 516-527.

51. Singh, Y. and A.S. Chauhan, NEURAL NETWORKS IN DATA MINING. Journal of Theoretical & Applied Information Technology, 2009. 5(1).

Figures and Tables

Table 1. Summary of Avatar interactions themes as per Beaudoin et al. 2021.

Avatar themes	**Examples**
Accusations	“You did this.“
Omnipotence	“I am all over the place.’’
Beliefs	“I think you are crazy.’’
Active listening, empathy	“Please relax, take your time. ’’
Incitements, orders	“You should stop doing.’’
Coping mechanisms	“Tell me why you are sad when I say this? ’’
Threats	“I will destroy you. ’’
Negative emotions	“It’s difficult for me to realize that.’’
Self-perceptions	“I identify myself as nothing.’’
Positive emotions	“I am the best in the world.’’
Provocation	“Try stopping me from making you ill. ’’
Reconciliation	“Should we make peace? ’’
Reinforcement	“Try this again. ’’

Table 2. Patient interactions’ themes as per Beaudoin et al. 2021.

Patient themes	**Examples**
Approbation	“You are right’’
Self-deprecation	“I can’t do this.’’
Self-appraisal	“I am a nice person.’’
Other beliefs	“You are the one controlling me’’
Counterattack	“You are the one who did this, not me!’’
Maliciousness of the voice	“You are trying to make this hard for all.’’
Negative	“It is very hard.’’
Negation	“I never did this.’’
Omnipotence	“I am the greatest.’’
Disappearance of the voice	“Please leave me alone!’’
Positive	’’I am feeling wonderful.’’
Prevention	“I am trying to dismiss you.’’
Reconciliation of the voice	“Can we work together?’’
Self-affirmation	“I am capable of doing this.’’

Table 3. Characteristics of sampled patients.

Characteristics	**Value (N=35)**
Sex (number of male, number of female)	27,8
Age (mean in years)	41.8 ± 11.2
Education (mean in years)	13.4 ± 3.2
Ethnicity (Caucasian, others)	94.3%,5.7%
% on Clozapine	45.7%

Table 4. Average performances of each classifier on the Avatar conceptual database for the metrics: accuracy, precision, recall and F1-Score

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-Score**
SVC	0.653680	0.736737	0.636364	0.636396
Linear SVC	0.705628	0.715403	0.675325	0.674928
Multinomial NB	0.437229	0.540432	0.545455	0.488000
Decision Tree	0.350649	0.403547	0.389610	0.388143
MLP	0.662338	0.658041	0.636364	0.636298

Table 5. Average performances of each classifier on the Patient conceptual database for the metrics: accuracy, precision, recall and F1-Score.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-Score**
SVC	0.526842	0.680169	0.526842	0.552448
Linear SVC	0.571930	0.610126	0.571930	0.575930
Multinomial NB	0.315789	0.529961	0.315789	0.297080
Decision Tree	0.350877	0.393063	0.350877	0.359419
MLP	0.564912	0.578114	0.564912	0.567399

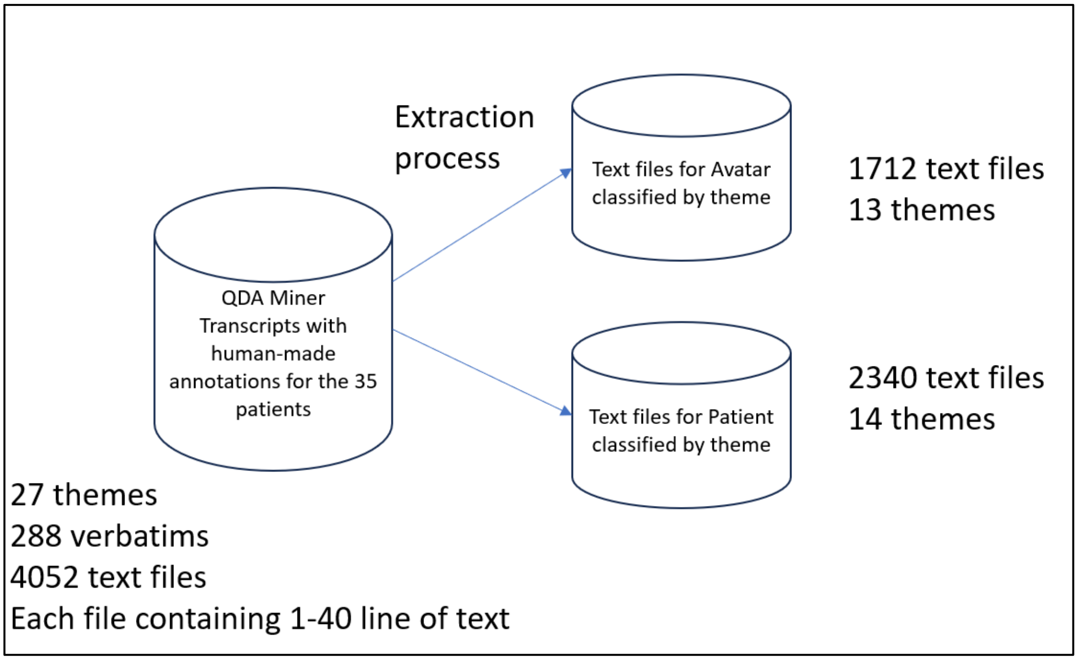
Extraction process
QDA Miner Transcripts with human-made annotations for the 35 patients
Text files for Avatar classified by theme
1712 text files
13 themes
Text files for Patient classified by theme
2340 text files
14 themes
27 themes
288 verbatims
4052 text files
Each file containing 1-40 line of text

Figure 1. Dataset for the corpus of Avatar Therapy

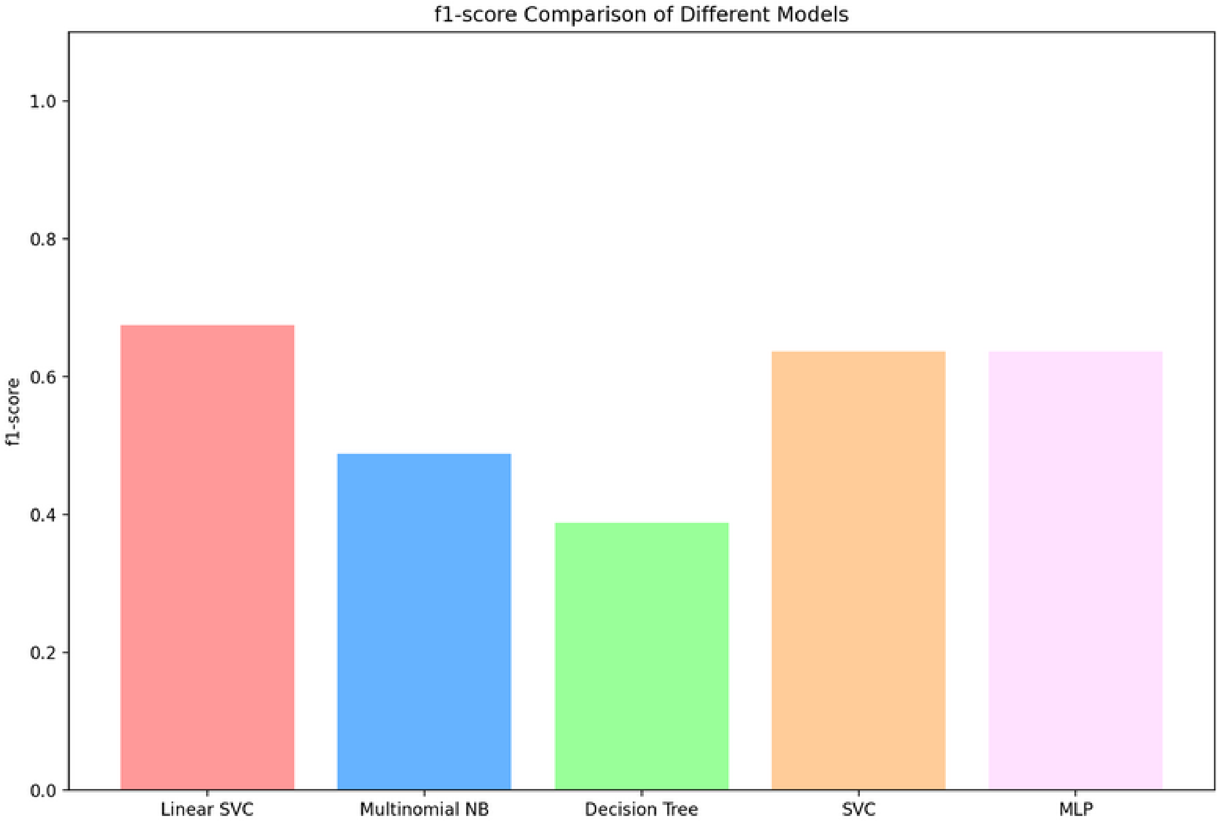

Figure 2. F1-Score comparisons of the different classifiers over the Avatar database.

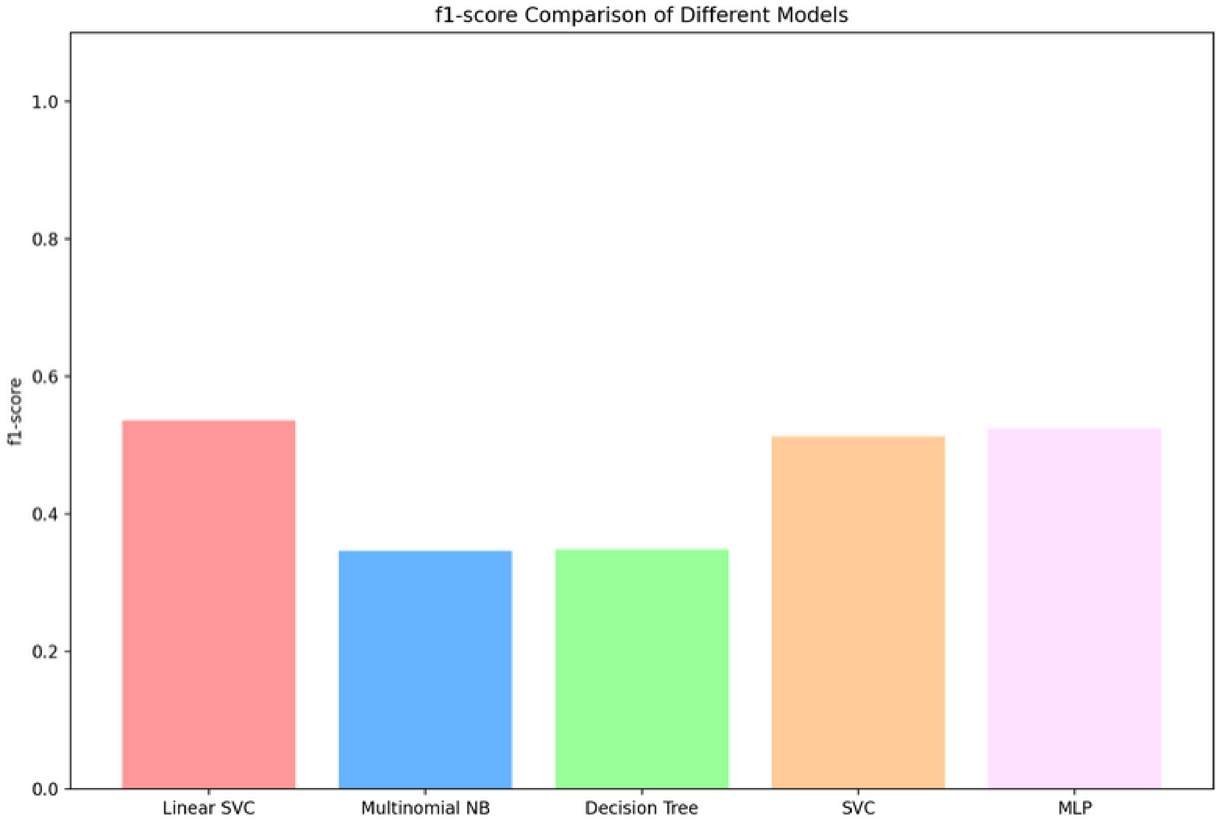

Figure 3. F1-Score comparisons of the different classifiers over the Patient database.

Article 5. Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach

Alexandre Hudon

Mélissa Beaudoin

Kingsada Phraxayavong

Laura Dellazizzo

Stéphane Potvin

Alexandre Dumais

Publié dans *Health Informatics Journal (2022).*

Abstract

Avatar Therapy (AT) is a modern therapeutic alternative for patients with schizophrenia suffering from persistent auditory verbal hallucinations. Its intrinsic therapeutical process is currently qualitatively analyzed via human coders that annotate session transcripts. This process is time and resource demanding. This creates a need to find potential algorithms that can operate on small datasets and perform such annotations. The first objective of this study is to conduct the automated text classification of interactions in AT and the second objective is to assess if this classification is comparable to the classification done by human coders. A Linear Support Vector Classifier was implemented to perform automated theme classifications on Avatar Therapy session transcripts with the use of a limited dataset with an accuracy of 66.02% and substantial classification agreement of 0.647. These results open the door to additional research such as predicting the outcome of a therapy.

Introduction

Psychotherapies imply complex social interactions that require the mobilisation of several cognitive and communication skills from both the patient and the therapist.1 Qualitative analysis of psychotherapy transcripts is often a methodology used to assess psychotherapies.2 However, this type of analysis often relies on human resources and remains rather time-consuming.3 Furthermore, qualitative approaches lack the generation of quantitative data to assess specific components of the intrinsic process of the psychotherapy.4 A growing body of researchers is attempting to use mixed methods to account for this problem. Consistency and coherence with the qualitative methodology employed is crucial for the process and is often infringed by the limits of subjectivity, notably when conducted by novice researchers.5 Inherent subjectivity biases from the researchers can also lead to issues in the validity and reliability of qualitative assessment of psychotherapeutic transcripts.6 These issues can be found in the annotations of in-person therapies, which are time-consuming, and the identification of the different interactions which can be even more complex. Annotations conducted using machine learning could be a potential solution to these issues to diminish this labor-extensive work and develop a systematic method to account for the potential inherent subjectivity biases of human annotators.7–8

Classification of textual entities is currently achieved in many different areas of medicine.9–11 Automated classification of text consists of analyzing a textual entity and classifying it under a specific label. This can be done by either supervised learning (i.e. an algorithm is trained with pre-existing data to conduct the classification) or unsupervised learning (i.e. labels are generated by the data).12 Text classification usually classifies text under two or more categories, which are also knowns as labels, features, or themes.13 The classification of therapeutic interactions may be a complex task as therapy sessions can vary in length, as well as content and sessions are dependent on the intrinsic and extrinsic characteristics of both therapist and patient.14 Few studies have attempted classify therapeutic interactions as large datasets consisting of human annotated transcripts, such as some seen in the field of internet-enabled cognitive behavioral therapy (IECBT), are required for complex machine learning algorithms to adequately learn and classify new information.15 However, in-person therapies can yield databases that are smaller than the ones generated by IECBT because of the need for human driven annotations which are time and resource demanding. This creates a need to find potential algorithms that can operate on small datasets. A recent systematic review having identified seven studies with small datasets in a psychotherapeutic context highlighted that support vector machine classifier was the best performing algorithm for these constraints.16 This opens the path for further studies on novel psychotherapeutic therapies for which limited data is available for analysis.

Avatar therapy (AT) is a type of virtual reality therapy. It is a modern therapeutic alternative for patients with schizophrenia suffering from persistent auditory verbal hallucinations (AVH) despite pharmacological treatment.17–19 Studies on Avatar Therapy taking place at our institution are currently analyzing the use of AT for patients diagnosed with schizophrenia with persistent auditory hallucinations and other mental illnesses. Patients currently enrolled in AT undergo nine weekly sessions of 45-min (one session to create the Avatar and 8 immersive sessions). An Avatar representing the most distressing voice of the patient is animated by the therapist to re-enact the voice in a secure therapeutic environment. The effects of AT on AVH are evaluated via the Psychotic Symptoms Rating Scale (PSYRATS total and PSYRATS-distress scores) and the Beliefs About Voices Questionnaire-Revised (BAVQ-R score) which are commonly used in the field to evaluate the effects of psychotherapy on schizophrenia patients. Other research teams such as Leff's and Craig's team in England are also using PSYRATS and BAVQ-R to assess AT.17,20 Current results demonstrate that therapeutic effects of AT on the distress associated with the voices were significant, as indicated by a net improvement in PSYRATS-distress score.21–22 In AT, the therapeutic process as a variable of effectiveness is of the upmost importance as there is an additional level of complexity added to the therapeutic dyad between the patient and the therapist, being the inclusion of an avatar. There are changes at a psychological level that are not captured by self-reported such as the PSYRATS and BAVQ-R. Traditional qualitative analyses consider these elements but have their own methodological limitations. The use of machine learning via text mining can be a complement to these analyses. Current attempts to evaluate the therapeutic processes of AT by the means of annotating interactions by themes has been entirely conducted by human evaluators.17,22,23 Furthermore, in AT, the complexity of having interactions between three individuals (avatar, therapist and the patient) and the fact that it is less readily available to the public limits the extent of useable data for constructing a dataset. The present study is therefore a first attempt at automated text annotation from a small dataset of AT transcripts.

The first objective of this study is to conduct the automated text classification of interactions in AT. Secondly, it is also important for us to assess if this classification is comparable to the classification done by human coders. This would provide an interesting solution for automated therapy annotations and could generate further data to evaluate AT process in relation to its effectiveness.

Methods

Dataset

A dataset was elaborated using 162 manually typed therapy transcripts of 18 randomly selected patients who undertook AT between 2017 and 2020 at our institution, which accounts for up to 10 therapy sessions per patient.23 The language of the transcripts was Canadian French. Transcripts were manually annotated using the 28 themes described in Beaudoin et al. 2021.

Please refer to Figure 1 in Beaudoin's study for classification of the themes. In the latter study, prior qualitative analysis of AT was conducted.23 Two research assistants coded each of the individual interactions independently. Robustness of the coding grid was cross validated by the same two research assistants. All the annotations were performed using QDA Miner version 5 (Provalis Research), a qualitative data analysis software. To improve the automated classification, annotations were then extracted as text files (containing from 1 to 40 interactions of the same theme) from QDA Miner and classified under three conceptual databases: Avatar, Patient and Therapist. The conceptual datasets were designed as per represented in Figure 1.

-- Please insert Figure 1 here--

Text files classification per theme from the qualitative analyses are represented in Table 1.

-- Please insert Table 1 here--

After reading from the database, the training sets for the Avatar, Patient and Therapist themes consisted of 691, 855 and 74 documents and the testing sets contained 231, 285, 32 documents respectively.

Machine learning algorithm

A support vector machine classifier was implemented to conduct the automated text classification (classify the different interactions under themes). Support vector machines encompass multiple algorithms that are often used in conjunction with tokenizers to evaluate the textual entities being classified. A tokenizer applies the process of tokenization, which is a method that breaks text into tokens to weight the value of a word or a sequence of words to compare it with other words or sentences.24 A member of the SVM family is the linear support vector classifier (LSVC). LSVC have been consistently more successful in text classification for small databases, such as ours.25 Prior review of algorithms for small datasets indicated that LSVC is the algorithm of choice for our study. LSVC was implemented using Python version 3.6.7 and Scikit-

Learn open library.26 It is noteworthy that Python was selected as the main programming language for our study because of its various uses in the domain of artificial intelligence, its flexibility as compared to other programming languages for scientific purposes and its support for many operating systems.27 Combined with a term frequency-inverse document frequency statistic (TF-IDF), it is an algorithm that performs best with text classification as compared to other combinations of SVM with a tokenizer.28 For the TDI-DF tokenization, the TfidfVectorizer offered in the Scitkit-Learn open library was selected as it enables to convert the raw text of the extracted interactions from the to-be annotated interview into numerical vectors. Vectorizers can be customized to account for stop-words. Considering the classification categories were designed in a way that text entities would be separated as per their intrinsic characteristics defined in Beaudoin et al. (2021) which are fundamentally different, the features are assumed to be linearly separable.23

To ensure best performances for the LSVC algorithm and enhance search strategies, a GridSearchCV (GSCV) was used. A GSCV is useful as it enables the user to test for different hyper-parameters and cross-validate the classification made by the LSVC to determine the best combination of LSVC parameters and the TfidVectorizer parameter variables.29

For each of the conceptual databases, LSVC has been trained using 70% of the available annotated documents and the remaining 30% has been used for testing purposes to establish a statistical probability (predictive score) that an interaction could be adequately classified. The training and testing sets did not overlap as per design recommendations.30 The predictive score refers to the mean accuracy (F1-Score) of the themes being testing. It is to be noted that a 70% training set and 30% testing set is the default setting for the Scikit-Learn LSVC library and is common practice for text classification31 This is modelized in Figure 2. A tenfold cross-validation was performed using the KFold model from the Scikit-Learn suite.

-- Please insert Figure 2 here--

The annotation process is shown in Figure 3. Each sentence in the transcript was regarded as an interaction.

-- Please insert Figure 3 here--

Performance analysis and inter-rater agreement

Information about the classification (precision, recall and F1-Score) for each theme was collected using the Classification Report tool, readily available in the Scikit-Learn metrics module. Precision refers to the positive predictive value, whereas recall refers to the sensitivity of the prediction and F1-score to the accuracy. The F1-score is the most widely used measure in text classification, reflecting the accuracy of theme classification and is a balance between precision and recall.32

While the F1-Score reflects the accuracy of theme classification, it does not account for the expected chance agreement. A Scott's Pi measure was therefore used to compare the degree of agreement between LSVC automatic classified annotation and the previously agreed ''correct'' annotation by human referees.33 The benchmark for the Scott's Pi measure interpretation tends to vary. The benchmark provided by the SAGE Research Methods was used in which a Scott's Pi of 0.81–1.00 is indicative of an almost perfect agreement, 0.61 to 0.80 of a substantial agreement, 0.41 to 0.60 of a moderate agreement, 0.21 to 0.40 of a fair agreement, 0.0–0.20 of a slight agreement and less than 0 as a poor agreement. This will be compared to the Scott's Pi agreement obtained between human annotators that was of 0.58 for our database.33

Results

The LSVC in combination with the TDI-DF was implemented and tested. An un-annotated transcript of an AT immersive session was automatically annotated. Training sets and testing sets are divided between Avatar themes (interactions involving the therapist animating the Avatar), Patient themes (patient's interactions) and Therapist theme (interactions involving the therapist talking directly to the patients).

The GSCV best selection of parameters for our study and our dataset indicated that document frequency and tolerance parameters are more important than others for our vectorizer and our

LSVC classifier. For our vectorizer, a minimum document frequency of 2 and maximum document frequency of 100 are applied. This ensures that a document appears at least 2 times to be considered by the LSVC and the limit of 100 is used to avoid documents that are repeated too frequently. The classifier tolerance was set to 0.001, dual parameters to false and intercept parameters to true. The mean squared error (MSE) training result was 0.88 and the MSE testing result was 0.96.

The Avatar, Patient and Therapist themes classification predictive score reached 70.6%, 61.8% and 100.0% respectively on average after 10 iterations. Considering the Therapist themes consists of solely one category, it was excluded from the overall weighed score. Therefore, an overall weighed score of 66.02% was obtained from the Avatar and Patient classifications. Classification reports in terms of precision, recall and F1-Score for Avatar, Patient and Therapist themes are listed in Table 2.

-- Please insert Table 2 here--

As it can be observed in Table 2, F1-Score for Avatar themes were on average better than for Patient's themes (0.706 vs 0.62). The theme Provocation performed the worst for the Avatar whereas Maliciousness of the voice performed the worst for Patient.

Agreement between human referees and the classifier reached a Scott's Pi of 0.647, which is ranked as substantial as per the SAGE Research Methods benchmark for Scott's Pi interpretation.

Discussion

The objective of this study was to conduct the automated text classification of interactions held during sessions of AT. This was conducted by implementing an LSVC algorithm.

It was possible to obtain a fully automated annotation of an un-annotated AT transcript. The weighed F1 predictive score for the annotation of the themes of the Avatar of 70.1% outperformed the F1-score for the themes of the Patient by 8.8%. The themes of the Avatar scored accuracies ranging from 54 to 94%. Regarding the themes of the Patient, the interaction theme Approbation (interactions in which patients completely or partially approved what their avatar was saying in response to a verbal attack) scored worst than all the other themes with a specificity of 15%. Since this theme contained 67 text files but scored less than themes with much less text files (e.g. Reconciliation of the Voice with 41) this may indicate that our conception of Approbation was perhaps not as distinct and might overlap with other themes. Therefore, a therapist evaluating the therapeutic process of the therapy with the same set of descriptive themes such as the 28 themes used in this study could reflect on this and revise the requirements for an interaction to be classified as Approbation. Considering that the latter is a coping mechanism distinct from other interactions held by the patient, we decided to keep it to explore whether this classification theme would need to be re-evaluated or would increase in homogeneity with additional data. In a similar study, in which medical reports that targeted patient symptoms were classified by severity, reports in their severe category versus their moderate category were often incorrectly classified because there are elements similar to both severe and moderate data that overlap in the definition of these two categories. 34 This supports the idea for better homogeneity amongst individual themes. Other Patient themes that had an overall poor accuracy included Self-deprecation (accuracy of 44% with 19 trained items and 8 tested items), Maliciousness of the voice (accuracy of 45% with 28 trained items and 12 tested items) and Omnipotence 12 (accuracy of 56% with 28 trained items and 12 tested items). This may be justified by the lack of data for these themes as compared to other accurately classified themes. These imbalances may also be explained by the fact that classifiers tend to respond better when there is a similar amount of data for each theme in the database.35 Fortunately, this gave us an initial insight on the therapeutic processes of the therapy as it outlined which interactions occurred potentially less during therapy sessions.

The secondary objective was to assess if this classification is comparable to the classification done by human coders.

Agreement between human referees and the classifier reached a Scott's Pi ranked as substantial. This is consistent with the Scott's Pi agreement of 0.588 between the two human referees having used QDA Miner. The fact that it was like the kappa agreement between human annotators indicates that both agreements for the annotation are comparable. For studies with datasets of similar size as the one used in this study (i.e. of less than 10 000 items) such as Balakrishnan, Zolnoori and Singh's studies reached substantial to moderate agreements.36–38 Although these agreements appear higher than ours, they employed a different base calculation than the Scott's Pi agreement. Pairwise agreement formula in Zolnoori's study is an improvement on Cohen's Kappa calculated at the level of entities rather than sentences to improve consistency and their result is also validated as substantial. This is also the case in Ewbanks et al.'s study in which they had 24 features and reached a kappa considered as moderate.15 As a comparison, a small study had 100 document samples and reached a kappa agreement of 0.6 with a Naïve Bayesian algorithm which was noted as acceptable.39

Limitations of this study include the classification F1-score for each theme that can be underestimated because of the 70% training and 30% testing set selection. It has been known to lead to class imbalances and sample representativeness issues.35 Such division in training and testing data is common for text classification, which is the reason we opted for such method. Nevertheless, since our dataset improves with additional data, it will be interesting to test for different training set sizes.31 It is to be noted that the transcripts analyzed in our study were typed in Canadian French and we did not find vectorizers that included stop-words (words not to be weighted during the process of tokenization as described above) for Canadian French language. This can yield a lower accuracy as there are insignificant words that can be weighed as part of a word vector.40

Conclusion

Machine learning can be beneficial to the field of psychiatry. Automated text classification for AT is a promising avenue to generate quantitative and qualitative data in an efficient way to be readily available to analyze. Our study allowed to automatically annotate an un-annotated transcript basing ourselves solely on a database derived from the transcripts of 18 patients. Reaching an agreement in the same range as human agreement, this study highlights that the task of annotation can be done by a machine, saving resources, which can improve the focus on patients' needs. It could also sharpen the therapeutic processes by reviewing what went wrong and what went well during AT based on automated text analysis. Nevertheless, this is to our knowledge the first study that outlined the possibility of automated annotation for AT and it highlights the need for more development in this field. These results open the door for additional research such as predicting the outcome of a therapy.

Ethical approval

This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Fondation Pinel, Chaire Eli Lilly Canada de recherche en schizophrénie, Services et recherches psychiatriques AD, Otsuka Canada Pharmaceutical and Le Fonds de recherche du Québec – Santé (FRQS).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Papageorgiou A, Loke Y, Fromage M. Communication skills training for mental health professionals working with people with severe mental illness. Cochrane Database of Systematic Reviews. 2017;2017(6).

2. Perepletchikova F. On the topic of treatment integrity. Clinical Psychology: Science and Practice. 2011;18(2):148-153.

3. Anderson C. Presenting and Evaluating Qualitative Research. American Journal of Pharmaceutical Education. 2010;74(8):141.

4. Szymańska A, Dobrenko K, Grzesiuk L. Characteristics and experience of the patient in psychotherapyand the psychotherapy's effectiveness. A structural approach. Psychiatria Polska. 2017;51(4):619-631.

5. Ranjbar M, Khankeh H, Khorasani-Zavareh D, Zargham-Boroujeni A, Johansson E. Challenges in conducting qualitative research in health: A conceptual paper. Iranian Journal of Nursing and Midwifery Research. 2015;20(6):635.

6. Noble H, Smith J. Issues of validity and reliability in qualitative research. Evidence Based Nursing. 2015; 18: 34–35.

7. Sebastiani F. Machine learning in automated text categorization. ACM Computing Surveys. 2002;34(1):1-47.

8. Ewbank M, Cummins R, Tablan V, Catarino A, Buchholz S, Blackwell A. Understanding the relationship between patient language and outcomes in internet-enabled cognitive behavioural therapy: A deep learning approach to automatic coding of session transcripts. Psychotherapy Research. 2020;31(3):300-312.

9. Wang Y, Sohn S, Liu S, Shen F, Wang L, Atkinson E et al. A clinical text classification paradigm using weak supervision and deep representation. BMC Medical Informatics and Decision Making. 2019;19(1).

10. García Adeva J, Pikatza Atxa J, Ubeda Carrillo M, Ansuategi Zengotitabengoa E. Automatic text classification to support systematic reviews in medicine. Expert Systems with Applications. 2014;41(4):1498-1508.

11. Venkataraman G, Pineda A, Bear Don't Walk IV O, Zehnder A, Ayyar S, Page R et al. FasTag: Automatic text classification of unstructured medical narratives. PLOS ONE. 2020;15(6):e0234647.

12. Spasic I, Nenadic G. Clinical Text Data in Machine Learning: Systematic Review. JMIR Medical Informatics. 2020;8(3):e17984.

13. Yao L, Mao C, Luo Y. Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. BMC Medical Informatics and Decision Making. 2019;19(S3).

14. Smink W, Sools A, van der Zwaan J, Wiegersma S, Veldkamp B, Westerhof G. Towards text mining therapeutic change: A systematic review of text-based methods for Therapeutic Change Process Research. PLOS ONE. 2019;14(12):e0225703.

15. Ewbank M, Cummins R, Tablan V, Bateup S, Catarino A, Martin A et al. Quantifying the Association Between Psychotherapy Content and Clinical Outcomes Using Deep Learning. JAMA Psychiatry. 2020;77(1):35.

16. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms. JMIR Mental Health. 2021;8(10):e22651.

17. Craig T, Rus-Calafell M, Ward T, Leff J, Huckvale M, Howarth E et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. The Lancet Psychiatry. 2018;5(1):31-40.

18. du Sert O, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophrenia Research. 2018;197:176-181.

19. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff A. Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. British Journal of Psychiatry. 2013;202(6):428-433.

20. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff A. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work?. Psychosis. 2013;6(2):166-176.

21. du Sert O, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophrenia

Research. 2018;197:176-181.

22. Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clinical Psychology & Psychotherapy. 2018;25(6):878-885.

23. Beaudoin M, Potvin S, Machalani A, Dellazizzo L, Bourguignon L, Phraxayavong K et al. The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clinical Psychology & Psychotherapy. 2021;

24. Ozaydin B, Zengul F, Oner N, Delen D. Text-mining analysis of mHealth research. mHealth. 2017;3:53-53.

25. Shridhar K, Dash A, Sahu A, Pihlgren G, Alonso P, Pondenkandath V et al. Subword Semantic Hashing for Intent Classification on Small Datasets. 2019 International Joint Conference on Neural Networks (IJCNN). 2019;

26. Hao J, Ho T. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. Journal of Educational and Behavioral Statistics. 2019;44(3):348-361.

27. Oliphant T. Python for Scientific Computing. Computing in Science & Engineering. 2007;9(3):10-20.

28. Busagala L, Ohyama W, Wakabayashi T, Kimura F. Multiple Feature-Classifier Combination in Automated Text Classification. 2012 10th IAPR International Workshop on Document Analysis Systems. 2012

29. Bisong E. More Supervised Machine Learning Techniques with Scikit-learn. Building Machine Learning and Deep Learning Models on Google Cloud Platform. 2019; 287-308.

30. Veronese E, Castellani U, Peruzzo D, Bellani M, Brambilla P. Machine Learning Approaches: From Theory to Application in Schizophrenia. Computational and Mathematical Methods in Medicine. 2013:1-12.

31. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation [Internet]. ScholarWorks@UTEP. 2022 [cited 2022 Jan 10];Available from: https://scholarworks.utep.edu/cs_techrep/1209/

32. Zhang D, Wang J, Zhao X. Estimating the Uncertainty of Average F1 Scores. Proceedings of the 2015 International Conference on The Theory of Information Retrieval. 2015

33. Allen M. Intercoder Reliability Techniques: Scott’s Pi. The SAGE Encyclopedia of Communication Research Methods. 2017; 1: 753-755

34. Karystianis G, Nevado A, Kim C, Dehghan A, Keane J, Nenadic G. Automatic mining of symptom severity from psychiatric evaluation notes. International Journal of Methods in Psychiatric Research. 2017;27(1):e1602.

35. Liu H, Cocea M. Semi-random partitioning of data into training and test sets in granular computing context. Granular Computing. 2017;2(4):357-386.

36. Balakrishnan V, Khan S, Arabnia H. Improving cyberbullying detection using Twitter users’ psychological features and machine learning. Computers & Security. 2020;90:101710.

37. Zolnoori M, Fung K, Patrick T, Fontelo P, Kharrazi H, Faiola A et al. A systematic approach for developing a corpus of patient reported adverse drug events: A case study for SSRI and SNRI medications. Journal of Biomedical Informatics. 2019;90:103091.

38. Singh V, Shrivastava U, Bouayad L, Padmanabhan B, Ialynytchev A, Schultz S. Machine learning for psychiatric patient triaging: an investigation of cascading classifiers. Journal of the American Medical Informatics Association. 2018;25(11):1481-1487.

39. de Ávila Berni G, Rabelo-da-Ponte F, Librenza-Garcia D, V. Boeira M, Kauer-Sant’Anna M, Cavalcante Passos I et al. Potential use of text classification tools as signatures of suicidal behavior: A proof-of-concept study using Virginia Woolf’s personal writings. PLOS ONE. 2018;13(10):e0204820.

40. Venkatasubramanian S, Veilumuthu A, Krishnamurthy A, C.E V, Nath K, Arvindam S. A non-syntactic approach for text sentiment classification with stopwords. Proceedings of the 20th international conference companion on World wide web - WWW '11. 2011;

Figures and Tables

Table 1. Distribution of text files per theme in the database

Avatar Themes	**Number of text files**	**Patient Themes**	**Number of text files**	**Therapist Theme**	**Number of text files**
Accusations	132	Approbation	67	Therapeutic intervention	106
Omnipotence	72	Self-deprecation	60		
Beliefs	89	Self-appraisal	87		
Active listening, Empathy	82	Other beliefs	88		
Incitements, Orders	48	Counterattack	86		
Coping mechanisms	82	Maliciousness of the voice	59		
Threats	31	Negative	129		
Negative emotions	49	Negation	90		
Self-perceptions	69	Omnipotence	67		
Positive emotions	43	Disappearance of the voice	80		
Provocation	87	Positive	81		
Reconciliation	60	Prevention	101		
Reinforcement	78	Reconciliation of the voice	41		
		Self-affirmation	104		
Total number of text files	922	*Total number of text files*	1140	*Total number of text files*	106
Average of text files per theme	71	*Average of text files per theme*	81	*Average of text files per theme*	106

Table 2. Precision, Recall and F1-Score for Avatar, Patient and Therapist themes.

Avatar Theme	**Examples (translated from French to**	**Precision (VPP)**	**Recall (Sensitivity)**	**F1-Score (Specificity)**	**Sample test size**

	English)				
Accusations	’’You did this’’	0.67	0.53	0.59	30
Omnipotence	’’I am the strongest’’	0.53	0.73	0.62	11
Beliefs	’’I believe that…’’	0.76	0.59	0.67	32
Active listening, Empathy	’’Take your time’’	0.76	0.8	0.78	20
Incitements, Orders	’’Kill yourself’’	0.67	0.91	0.77	11
Coping mechanisms	’’I am not happy when you say this’’	1	0.75	0.86	16
Threats	’’I will hurt you’’	1	0.91	0.95	11
Negative emotions	’’It’s difficult’’	0.72	0.87	0.79	15
Self-perceptions	’’The way I see myself is…’’	0.65	0.65	0.65	23
Positive emotions	’’I’m fine’’	0.9	0.6	0.72	15
Provocation	’’What are you waiting for?	0.43	0.71	0.54	14
Reconciliation	’’Should we make peace?’’	0.73	0.73	0.73	15
Reinforcement	’’You did well’’	0.7	0.78	0.74	18
	Average scores	0.73	0.71	0.706	231
Patient Themes	Examples	Precision (VPP)	Recall (Sensitivity)	F1-Score (Specificity)	Sample test size
Approbation	’’I agree with you’’	0.15	0.14	0.15	14
Self-deprecation	’’I could never be confident’’	0.32	0.75	0.44	8
Self-appraisal	’’I am kind’’	0.65	0.6	0.63	25
Other beliefs	’’You are controlling me’’	0.62	0.58	0.6	26
Counterattack	’’I think you are wrong’’	0.5	0.62	0.56	16
Maliciousness of the voice	’’You are spreading misfortune to all’’	0.5	0.42	0.45	12
Negative	’’It’s difficult’’	0.6	0.58	0.59	31
Negation	’’I do not recognize this’’	0.95	0.56	0.7	34
Omnipotence	’’I am the best’’	0.54	0.58	0.56	12

Disappearance of the voice	’’Go away’’	0.83	0.76	0.79	25
Positive	’’I’m fine’’	0.71	0.88	0.79	17
Prevention	’’I will try to ignore you’’	0.75	0.75	0.75	32
Reconciliation of the voice	’’I want to learn to live with you’’	0.55	0.75	0.63	8
Self-affirmation	’’I do not think so’’	0.58	0.6	0.59	25
Average scores		0.65	0.65	0.62	285
Therapist Theme	Examples	Precision (VPP)	Recall (Sensitivity)	F1-Score (Specificity)	Sample test size
Therapeutic intervention	Any intervention by the therapist	1	1	1	32
Average scores		1	1	1	32

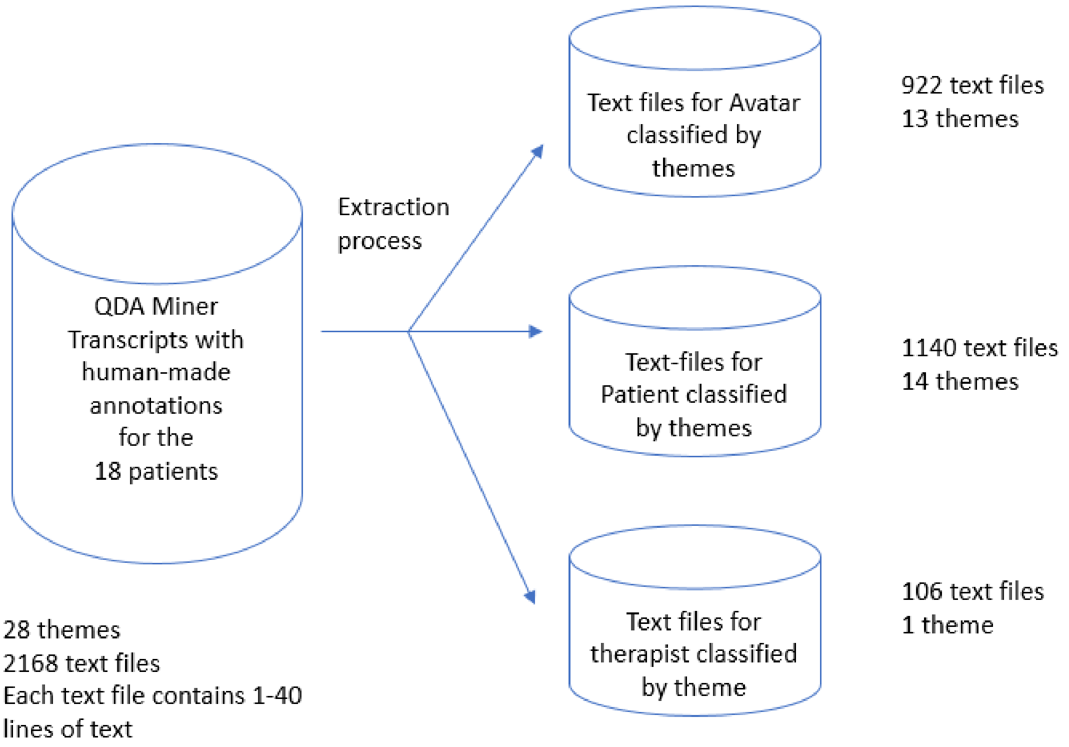
QDA Miner Transcripts with human-made annotations for the 18 patients
28 themes
2168 text files
Each text file contains 1-40 lines of text
Extraction process
Text files for Avatar classified by themes
922 text files
13 themes
Text-files for Patient classified by themes
1140 text files
14 themes
Text files for therapist classified by theme
106 text files
1 theme

Figure 1. Conceptual datasets design

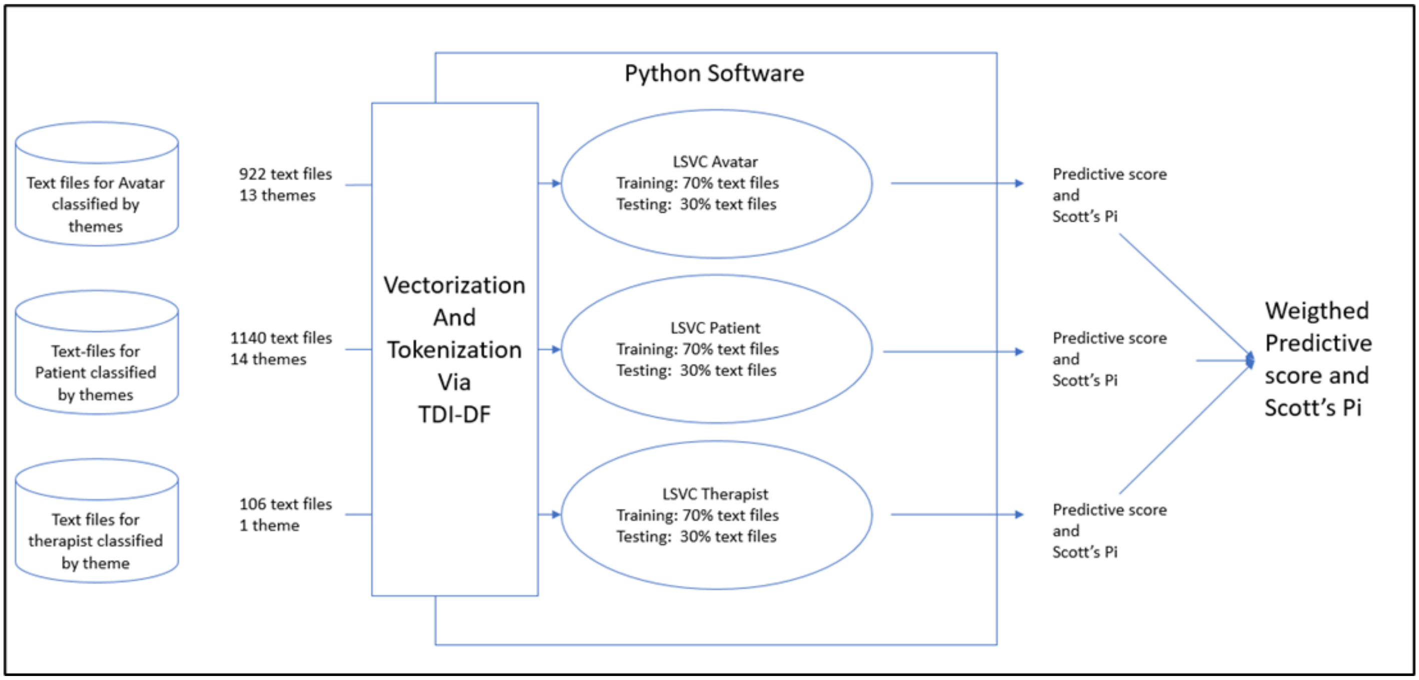
Python Software
Text files for Avatar classified by themes
Text-files for Patient classified by themes
Text files for therapist classified by theme
922 text files
13 themes
1140 text files
14 themes
106 text files
1 theme
Vectorization And Tokenization Via TDI-DF
LSVC Avatar
Training: 70% text files
Testing: 30% text files
LSVC Patient
Training: 70% text files
Testing: 30% text files
LSVC Therapist
Training: 70% text files
Testing: 30% text files
Predictive score and Scott's Pi
Predictive score and Scott's Pi
Predictive score and Scott's Pi
Weigthed Predictive score and Scott's Pi

Figure 2. Implementation of LSVC on conceptual databases to derive a Predictive score and a Scott's Pi.

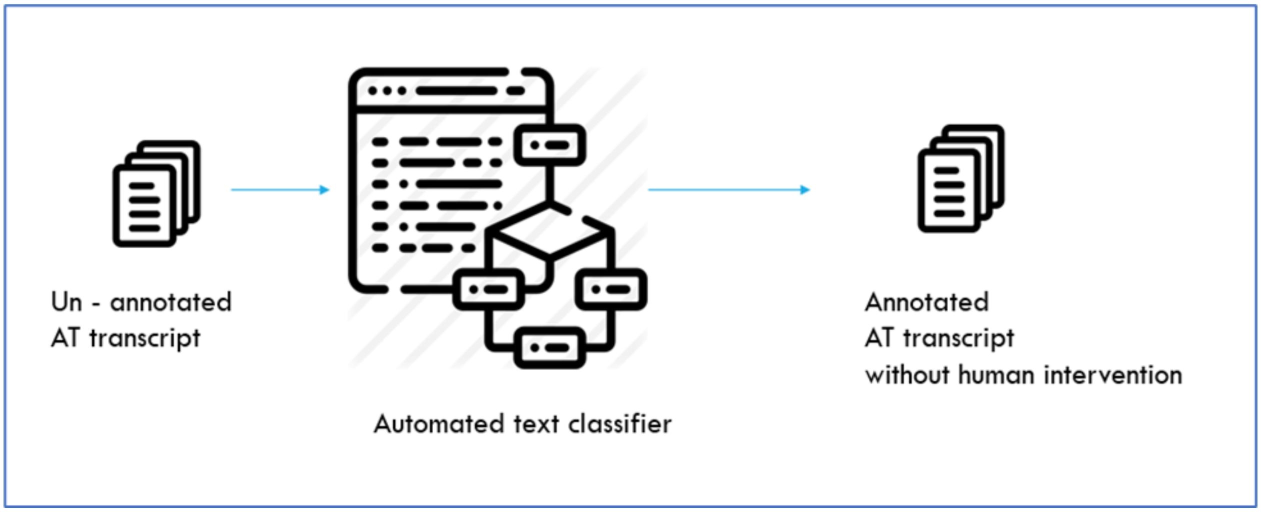

Figure 3. Annotation process overview

Article 6. Unsupervised Machine Learning Driven Analysis of Verbatims of Treatment-Resistant Schizophrenia Patients Having Followed Avatar Therapy

Alexandre Hudon

Mélissa Beaudoin

Kingsada Phraxayavong

Stéphane Potvin

Alexandre Dumais

Publié dans *Journal of Personalized Medicine (2023).*
Article révisé par les pairs.

Abstract

(1) Background: The therapeutic mechanisms underlying psychotherapeutic interventions for individuals with treatment-resistant schizophrenia are mostly unknown. One of these treatment techniques is avatar therapy (AT), in which the patient engages in immersive sessions while interacting with an avatar representing their primary persistent auditory verbal hallucination. The aim of this study was to conduct an unsupervised machine-learning analysis of verbatims of treatment-resistant schizophrenia patients that have followed AT. The second aim of the study was to compare the data clusters obtained from the unsupervised machine-learning analysis with previously conducted qualitative analysis. (2) Methods: A k-means algorithm was performed over the immersive-session verbatims of 18 patients suffering from treatment-resistant schizophrenia who followed AT to cluster interactions of the avatar and the patient. Data were pre-processed using vectorization and data reduction. (3): Results: Three clusters of interactions were identified for the avatar's interactions whereas four clusters were identified for the patient's interactions. (4) Conclusion: This study was the first attempt to conduct unsupervised machine learning on AT and provided a quantitative insight into the inner interactions that take place during immersive sessions. The use of unsupervised machine learning could yield a better understanding of the type of interactions that take place in AT and their clinical implications.

Keywords

psychotherapy; virtual-reality therapy; auditory hallucinations; schizophrenia; machine learning

Introduction

Schizophrenia is a severe mental illness that affects millions of people worldwide and can profoundly impact the affected individual, their families, and society as a whole [1,2,3]. Chronic psychotic symptoms can make it difficult for individuals with the illness to maintain relationships, hold a job, and have a fulfilling life [3]. Moreover, living with schizophrenia leads to a significantly reduced life expectancy due to a much higher risk of completing suicide and suffering from chronic physical conditions such as cardiovascular diseases or diabetes [4,5]. The societal burden of this illness is quite high given the loss of productivity and the substantial costs associated with treating schizophrenia (i.e., hospitalizations, regular healthcare appointments, and medications) [2,6]. Most distressing acute symptoms can be substantially reduced using antipsychotic medications; however, up to a third of patients fail to improve, making them resistant to treatment [7]. These individuals often have a poorer quality of life, experience more frequent hospitalizations, have higher rates of suicide, leading to significantly higher societal costs compared to those who respond appropriately to antipsychotics [8]. The most effective medication for this condition, clozapine, is not always an option since it has poor tolerability and requires careful monitoring for severe side effects [9,10]. Moreover, a significant subset of treatment-resistant patients also fails to respond to clozapine; these are often referred to as being “ultra-resistant” to treatment [11].

To learn how to cope with their persistent symptoms, patients with treatment-resistant schizophrenia are generally encouraged to undergo psychotherapy [12]. The most prevalent and distressing symptom is auditory verbal hallucinations (AVH) i.e., hearing voices); therefore, this specific component of schizophrenia is targeted by a few psychotherapeutic approaches [13,14]. The most studied and widespread one is cognitive–behavioral therapy, which has been shown to be significantly more effective than a control condition in reducing the frequency and distress associated with AVH in this population [15]. However, the effect size is only moderate and the symptoms of only a small subset of patients are reduced in a clinically significant manner [16,17,18,19]. Additionally, according to a recent meta-analysis, CBT for psychosis might have little to no impact on quality of life [19]. This could be due to the fact that this therapy, largely based on psychoeducation and mindfulness, does not offer the patient an opportunity to practice interacting with their voices and finding new coping strategies under therapeutic supervision. To address this gap, a few novel therapeutic approaches are now focused on having the patient improve their relationship with their voice(s), notably by entering a dialogue with them [20]. This can be achieved using different techniques such as chairwork (i.e., having the patient take the role of the voice in one chair and then answering them in a different chair), through role-play with the therapist, or by dialoguing with an avatar representing the distressing voice [20]. Avatar therapy (AT), which was initially developed using an avatar on a 2D screen, has now been adapted to virtual reality (VR), thereby increasing the immersive aspect of psychotherapy [18,21,22]. In this therapy, patients with treatment-resistant schizophrenia are first invited to create and personalize an avatar resembling the mental image that they have of their most distressing hallucination, both in terms of physical appearance and tone of voice. Afterward, patients undergo six to ten one-hour weekly therapeutic sessions which all include approximately 5 to 20 min of dialogue with their avatar in VR. The avatar is animated by the therapist, who is installed in a separate room and has their voice modified in real-time. In addition to role-playing the voice, the therapist also has control over the facial expressions as well as the distance between the avatar and the patient. During the first few sessions, the therapist starts by repeating verbatim what the patient reports that their voice usually says, and mostly uses provocative techniques. For example, the therapist, animating the avatar, might repeat “you are worthless”. However, the avatar gradually opens to the patient and starts using more and more positive techniques [18,21,22]. The different themes addressed during AT have been described in detail in a previous qualitative study by Beaudoin and her team [23]. Notably, the avatar mainly used techniques that were classified as provocative (e.g., threats, accusations, affirmations of omnipotence) or positive (e.g., reinforcement, empathetic listening). The patients responded in a few different ways: with an emotional response (positive, neutral, or negative), by mentioning beliefs about the voices and/or schizophrenia (e.g., omnipotence, malevolence), self-perceptions (i.e., self-appraisal or self-deprecation), coping mechanisms (e.g., self-affirmation, counterattack), or aspirations (e.g., prevention strategies) [23,24].

While previous qualitative studies highlight promising avenues to better comprehend the inner psychotherapeutic processes that might be linked to a positive outcome, it is possible that some elements are underexamined or prone to subjective biases, which are prevalent in such studies [25]. The use of artificial-intelligence-driven approaches, such as unsupervised machine learning, is an increasingly seen technique in various medical fields in order to derive objective data from several types of textual datasets (and other sources of datasets) in the medical field [26]. It is a technique in which unlabeled data are used to conduct different types of tasks such as hierarchical learning, data clustering, latent variable modeling, dimensionality reduction (on large datasets), and outlier detection [27]. A few implementations of such algorithms are found in psychiatry. For example, recent research conducted by Kung et al. (2022) used unsupervised learning to identify qualitative subtypes of depression based on the clinical data from 18,314 patients with depression [28]. Another recent example is the identification of five subgroups of psychosis amongst 765 individuals suffering from DSM-IV diagnoses of schizophrenia, bipolar affective disorder (I/II), schizoaffective disorder, schizophreniform disorder, and brief psychotic disorder by using clustering methods: affective, suicidal, high functioning, depressive, and severe psychosis [29]. In the field of psychotherapy and psychotherapeutic approaches, the latest literature review on the subjective identifies nine studies that used unsupervised machine learning [30]. Most of these applications were used to perform human-like responses to interact with patients after learning from datasets of multiple interactions derived from thousands of therapy sessions. An example of such application is the development of ClientBot, by Tanana et al. (2019), which used natural language-processing methods for automated coding rather than human coders to perform interactions with the patients [31]. To our knowledge, the use of unsupervised machine learning to objectively assess verbatims from AT has never been conducted. Natural-language-processing (a subset of machine learning) approaches for patients suffering from schizophrenia are currently being studied and demonstrate promising avenues to characterize sub-clinical linguistic differences in schizophrenia-spectrum disorders which might be clinically relevant [32]. Analysis of verbatims using unsupervised learning might therefore provide insights as to different types of interactions taking place during the immersive sessions.

This study's primary aim was to conduct an unsupervised machine-learning analysis of verbatims of treatment-resistant schizophrenia patients that had followed AT. The second aim of the study was to compare the data clusters obtained by the unsupervised machine-learning analysis with the main themes identified by Beaudoin et al. (2021) through human-driven qualitative analysis. The hypothesis was that unsupervised machine-learning analysis will provide clusters similar to the main themes identified by Beaudoin et al., while providing insight as to how certain themes might be sub-divided.

Materials and Methods

Participants and Recruitment

The participants included in this study received AT as part of pilot trials at the Centre de recherche de l'Institut universitaire en santé mentale de Montréal (CR-IUSMM) and one ongoing trial comparing AT to CBT. The participants all belonged to the clinical trial registered on Clinicaltrials.gov (identifier number: NCT03585127) [18,21]. Included participants received nine psychotherapeutic sessions, of which eight were immersive. In these sessions, the patients interacted with an avatar representing their most significant AVH. The participants included in this study were all patients at the IUSMM, over 18 years of age, who were suffering from treatment-resistant schizophrenia as defined by the absence of response to two or more antipsychotics, and who had received AT between 2017 and 2020. The ethics committee of CR-IUSMM approved the study as part of the protocol for AT.

Data collection

First, a content analysis was performed on 125 immersive sessions (1419 min of therapy) from 18 patients with treatment-resistant schizophrenia who underwent avatar therapy in the context of either one of two clinical trials assessing the efficacy of this therapy [24]. Each therapy session was first transcribed (Canadian French), and then read and carefully annotated by each member of the research team. Discussions then took place every week to identify each theme and organize them hierarchically into a grid. Then, each verbatim (i.e., a group of sentences representing one expressed idea) was coded into one of the identified themes. Transcripts were annotated sequentially, and the grid was adjusted as the coding progressed in a back-and-forth manner, and the process only stopped when data saturation occurred (i.e., when the therapies of a few participants were coded without having to adjust the grid). To assess potential inter-rater variability, 63% of the sample was also coded by a second person; overall, the inter-rater agreement was fair for the detailed theme grid (Scott's Pi = 0.514) and moderate for agreement on the key themes only (Scott's Pi = 0.660). More details about the methodology and the results of this analysis can be found in a previous paper published on that matter [24].

From the above content analysis, two datasets were developed. The first dataset contained all the labeled interactions for the avatar and the patient, whereas the second dataset contained unlabeled interactions for the avatar and the patient as per Figure 1. In the labeled dataset, Beaudoin et al. (2021) identified two major categories of interactions for the avatar: confrontational techniques and positive techniques. For the patient, they identified five categories: self-perceptions, aspirations, emotional responses, coping mechanisms, and beliefs about voices and schizophrenia.

-- Please insert Figure 1 here--

Data Analysis

The various steps included in the data analysis are presented below. The overall flow of the data analysis process is presented in Figure 2.

-- Please insert Figure 2 here--

Unsupervised Machine-Learning Algorithm

A k-means clustering algorithm was used to conduct the clustering of the data from the dataset containing the unlabeled interactions using Python 3.9 with the Scitkit-Learn open library [33,34]. This widely used algorithm attempts to cluster similar data in an easy-to-interpret, relatively fast, and scalable way while guaranteeing convergence of the data [35]. It determines whether two items are identical and clusters them based on their Euclidean distance, representing the length of a line traced between two data points [36]. The number of clusters is determined in advance, and the following steps are performed iteratively [37]. First, the center of each cluster (centroid) is randomly selected, then the Euclidean distance of all data points to the centroids is calculated, and the data points are then assigned to the closest cluster. Then, the new centroid of each cluster is identified by taking the mean of all data points in the cluster and repeating the process until all the points converge and the cluster centers stop moving.

To determine the number of clusters, an elbow plot is used. This technique illustrates the global dissimilarity (also known as inertia) between the data points and the number of potential clusters. Dissimilarity refers to the squared Euclidean distance between the data points and the cluster centers, and global dissimilarity is, therefore, the sum of dissimilarity for all the data points within all the clusters [38]. The use of an elbow plot as compared to other techniques (i.e., Silhouette's coefficient and the gap statistic) was because of the smaller size of the dataset, the notion that the data might not be clearly separated, and the gain in time complexity [39].

Data Preprocessing

The term frequency-inverse document frequency statistic (TF-IDF) was used to convert the raw text of all the textual interactions into numerical vectors to be used by the k-means algorithm. Therefore, all the sentences of each text file included as part of the dataset are converted into a vector. This step is necessary because length between raw textual data points cannot be measured and compared [40]. Considering the wide variety of interactions that are taking place in AT and the previous knowledge of the qualitative insights of these interactions, these textual interactions were assumed to be linearly separable. The TfidfVectorized of the Scitkit-Learn open library was used [33].

Data Reduction

A principal component analysis (PCA) using the Scitkit-Learn open library was conducted on the vectorized data prior to the k-means analysis [33]. Reducing the dimensionality of a dataset is a method performed to increase interpretability while minimizing information loss [41]. PCA is among the most used algorithms for such tasks as it attempts to estimate the linear combinations of the different independent variables by creating uncorrelated variables (principal components) that will successively maximize variance. It accomplishes this by locating a collection of orthogonal vectors known as the principal components, which reflect the directions in which the data's largest variation occurs. Each consecutive principal component is chosen to be orthogonal to the previous ones and to capture the next biggest amount of variance. The first principal component corresponds to the direction with the largest amount of variation. Fewer dimensions and frequently easier analysis and visualization characterize the resulting converted dataset.

Comparing the Unsupervised Machine-Learning Clustering with the Labeled Data

A descriptive statistical analysis of the comparison between the previously labeled data and the clustered labeled data was performed. This was done using a simple Python 3.9 program that remapped all the unlabeled interactions from the unlabeled dataset with their labeled counterpart while keeping track of their newly identified cluster. As per Beaudoin et al. (2021), the frequency of each sub-theme was compared between both datasets.

Results

Vectorization and data reduction were successfully conducted for all the data points of the unlabeled dataset prior to performing clustering. Interactions from 922 text files were identified for the avatar and 1140 text files for the patient.

Clustering

It can be observed in Figure 3 that the avatar elbow curve indicates that the optimal number of clusters should be between two and four. Therefore, three clusters were selected as the initiation parameter for the k-means algorithm.

-- Please insert Figure 3 here--

As displayed in Figure 4, data points were scattered across the three different clusters. The red cluster appeared to have more homogeneous data points, whereas the blue cluster had data that were very far apart and more heterogeneous. In the middle of the graph, there appeared to be no clear delimitation across the three clusters which might have indicated that these data points were not clearly divisible into different clusters. These interactions could likely be susceptible to various diverging interpretations if they were to be qualitatively assessed by human coders.

-- Please insert Figure 4 here--

Examples of verbatims from the different clusters can be found below (translated from Canadian French to English):

Blue cluster:
“You are supposed to let me win.”
“They are right, you are the one that stole it.”
“I don’t believe you; you can’t be right.”

Green cluster:
“Do you believe in yourself?”
“Maybe you are becoming crazy? Are you?”
“Let’s make peace.”

Red cluster:
“How will you do it? What is it that you will do?”
“Do you want me to stay? Should I leave?”
“What could they do for you?”

As depicted in Figure 5, the patient elbow curve indicated that the optimal number of clusters should be around four. Therefore, four clusters were selected as the initiation parameter for the k-means algorithm.

-- Please insert Figure 5 here--

As displayed in Figure 6, data points were scattered across the four different clusters. The yellow and green clusters appeared to overlap, whereas the red and blue clusters were well delimited from all the other clusters. This indicated that some interactions clearly belonged together, whereas it was difficult to discriminate between interactions belonging to the yellow and the green clusters. The green cluster had very homogenous interactions, whereas the blue and red clusters were heterogeneous.

Examples of verbatims from the different clusters can be found below.

Blue cluster:

"I have weaknesses."

"You are right, I need to call my mother. It is important that I call her very soon."

"Yes, it is a fact. I'm not a good person."

Yellow cluster:

"I'd like you to stop talking to me."

"No, you can't. You are not allowed to do this to me."

"I would like you to give me positive energy and please stop trying to destroy me all the time."

Green cluster:

"Life is great, my friend."

"You are not so much in my head anymore."

"This week you left me alone. I like that."

Red cluster:

"I'll confront you and tell you that you are very ill."

"I need to stop playing slot machines."

"The doctor is helping me. He is my ally. He is telling me what to do."

Comparison with Previously Labeled Data

Cross-labeling of the unlabeled dataset with the labeled dataset was conducted. Table 1 presents the original division of the text files and their classification (labels) for the labeled dataset, whereas Table 2 displays textual entities' mapping and their classification per the unlabeled dataset. Compared to Beaudoin et al. (2021), the clustering analysis identified three clusters (labels) for the avatar interactions and four clusters for the patient interactions.

-- Please insert Table 1 here--

-- Please insert Table 2 here--

With the mapping of the labels on the clustering data, it can be observed for the avatar that some of the confrontational techniques appear to have been shared across the blue and green clusters. In contrast, the positive techniques were mostly spread across the green and the red clusters. Clustering highlights the heterogeneity of the interaction across these categories previously defined as confrontational techniques and positive techniques.

As for the patient, most of the interactions previously defined as per the five labels appear to have been clustered into the green and yellow clusters, especially emotional responses in the yellow cluster. They mostly regroup interactions that were previously classified as coping mechanisms, aspirations, and beliefs about voices and schizophrenia. The blue and red clusters appear to regroup interactions that were mainly scattered across the previously defined labels. Interactions previously labeled as coping mechanisms appear to be less present in the blue cluster, whereas they were more prevalent in the red cluster. The opposite classification can be observed with the interactions previously labeled as aspirations.

Discussion

The main goal of this study was to conduct unsupervised machine-learning analysis verbatims of treatment-resistant schizophrenia patients that had followed AT. This was done by vectorizing textual interactions of the avatar and the patient during immersive sessions of AT, reducing the complexity of the data, and performing a cluster classification of unlabeled data. That enabled the identification of three clusters for the avatar's interactions and four clusters for the patient's interactions. These unlabeled clustered data were then remapped as per the previous qualitative study on the same verbatims Beaudoin et al.

It was possible to observe three distinct clusters for the avatar interactions. Considering the variety of potential interactions that the therapist must employ during the immersive session to personalize the experience for each patient, it is possible that this provides a distinction between the confrontational techniques classified in the blue cluster and those in the red cluster. As indicated in O'Brien et al. (2021), in AT, the therapist must consider a formulation to inform the direction of the therapy as well as quickly responding as the characterized avatar [42]. Several studies outline the use of direct confrontation in psychotherapy as well as empathetic confrontation. Empathic confrontation is often observed as part of schema therapy to address patients' maladaptive behaviors and it also serves emotional regulation [43]. On the other hand, direct confrontation can be seen in AT to provoke the patient by mimicking their experience with AVH. Since these two techniques differ in terms of interactions and delivery, this might explain the division of these interactions between two clusters. As for positive techniques, they were also scattered across two clusters (red and green). A previous study on integrative psychotherapy indicated that therapeutic alliance had the most evidence as a predictor of patient change [44]. One challenge of AT is, therefore, to bring forward the personification of the AVH while maintaining the therapeutic alliance and inducing positive changes, which may imply different types of positive techniques. In CBT, psychotherapeutic approaches for patients with schizophrenia include the development of trust, normalization, coping strategy enhancement, and reality testing. In the green cluster, some of the interactions previously classified as positive techniques appear to partly include elements of confrontational techniques. That might be part of reality testing, which might appear confrontational while potentially assessing the self-perceptions or beliefs of the patient about their AVH.

The patient interactions were classified into four different clusters, meaning that the interactions might have been less heterogenous than what was found in previous qualitative study on the same dataset [23]. The blue cluster contained very few interactions, which suggests there were outliers in the interactions between the patients and their avatars. A recent study assessing 499 language samples with a natural language processing algorithm on patients with schizophrenia or bipolar disorder outlined that sociodemographic and individual differences should be considered while conducting language analysis for psychosis [45]. These, as well as relationships with others, were not specifically captured with the previously conducted qualitative analysis. That might account for the outliers identified in the blue cluster, but further analyses should be conducted. Most of the coping mechanism interactions were found in the green cluster, whereas the emotional responses were found in the yellow cluster while these two clusters seemed to intersect. That is not surprising considering that coping mechanisms and emotional responses are two strong components of psychotherapeutic approaches and are often tied together, considering that coping mechanisms involuntarily manifest when strong emotions are involved [46]. The overlap of these two clusters might therefore indicate that interactions reflecting coping mechanisms could be further detailed as per other characteristics. For example, coping mechanisms found in the green cluster might be more tied to aspirations and beliefs about voices and schizophrenia, whereas the ones found in the yellow cluster might be more tied to emotional responses.

Limitations

While using the k-means algorithm enabled clustering, a larger dataset would have been preferred to account for the errors linked to the centroids of the clusters being dragged by interactions that are outliers. It should be noted that the transcripts examined in our research were typed in Canadian French, and locating vectorizers which included stopwords was not possible to for that language. As insignificant terms can be considered part of a word vector, the accuracy may have been impacted. Another limitation was the small sample of patients involved in the presented study as it affected the generalizability of the study considering that the interactions identified were part of a small number of participants.

Conclusions

Unsupervised machine learning can be a beneficial approach in the mental health field, bringing an objective evaluation of verbatims of AT. Our study allowed the identification of three major clusters of interactions for the avatar's interactions and four major clusters for the patient's interactions. As compared to the previously established qualitative analysis realized by human coders on the same dataset, it was observed that the results for the clustering of avatar interactions were similar to the ones identified by human coders. However, there was a greater divergence for the patient interactions, which were scattered across the identified clusters. The interactions previously labeled coping mechanisms and aspirations were the two types that were mainly classified together, whereas the other labels were more heterogeneously scatted across the four clusters. This study was the first attempt to conduct unsupervised machine learning on AT and provides quantitative insight into the inner interactions taking place during immersive sessions. The consideration of further data, such as the addition of emotions or psychomotor indices, could be beneficial to better comprehend the inner processes of AT and evaluate its implication in regard to the therapeutic outcome.

Author Contributions

Conceptualization, A.H., M.B., K.P., S.P. and A.D.; methodology, A.H., M.B. and A.D.; validation, A.H. and A.D.; formal analysis, A.H.; investigation, A.H.; data curation, A.H. and M.B.; writing—original draft preparation, A.H.; writing—review and editing, A.H., M.B., S.P. and A.D.; supervision, K.P., S.P. and A.H.; project administration, K.P.; funding acquisition, K.P., S.P. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work was indirectly supported by Le Fonds de recherche du Québec—Santé (FRQS), Otsuka Canada Pharmaceutical Inc., Chaire Eli Lilly Canada de recherche en schizophrénie, MEI (Ministère de l'Économie et de l'Innovation), Services et recherches psychiatriques AD, and Fonds d’excellence en recherche Apogée Canada (Institut de la Valorisation des Données IVADO).

Institutional Review Board Statement

This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients. Participants were selected based on the proof-of-concept trials of Percy du Sert et al. (2018) and Dellazizzo et al. (2021) [18,21]. The trial was conducted in accordance with the Declaration of Helsinki and was approved by the institutional ethical committee (CER IPPM 16-17-06).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to patients’ privacy.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Deng SY, Wang YZ, Peng MM, Zhang TM, Li M, Luo W, et al. Quality of life among family caregivers of people with schizophrenia in rural China. Qual Life Res. 2023.

2. Goeree R, O'Brien BJ, Goering P, Blackhouse G, Agro K, Rhodes A, et al. The economic burden of schizophrenia in Canada. Can J Psychiatry. 1999;44(5):464-72.

3. Dong M, Lu L, Zhang L, Zhang YS, Ng CH, Ungvari GS, et al. Quality of Life in Schizophrenia: A Meta-Analysis of Com-parative Studies. Psychiatr Q. 2019;90(3):519-32.

4. Saha S, Chant D, McGrath J. A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time? Arch Gen Psychiatry. 2007;64(10):1123-31.

5. Ali S, Santomauro D, Ferrari AJ, Charlson F. Schizophrenia as a risk factor for cardiovascular and metabolic health out-comes: a comparative risk assessment. Epidemiol Psychiatr Sci. 2023;32:e8.

6. Kotzeva A, Mittal D, Desai S, Judge D, Samanta K. Socioeconomic burden of schizophrenia: a targeted literature review of types of costs and associated drivers across 10 countries. J Med Econ. 2023;26(1):70-83.

7. Millgate E, Hide O, Lawrie SM, Murray RM, MacCabe JH, Kravariti E. Neuropsychological differences between treat-ment-resistant and treatment-responsive schizophrenia: a meta-analysis. Psychol Med. 2022;52(1):1-13.

8. Kennedy JL, Altar CA, Taylor DL, Degtiar I, Hornberger JC. The social and economic burden of treatment-resistant schiz-ophrenia: a systematic literature review. Int Clin Psychopharmacol. 2014;29(2):63-76.

9. Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, et al. Comparative efficacy and tolerability of 15 antipsy-chotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet. 2013;382(9896):951-62.

10. Mortimer AM, Singh P, Shepherd CJ, Puthiryackal J. Clozapine for treatment-resistant schizophrenia: National Institute of Clinical Excellence (NICE) guidance in the real world. Clin Schizophr Relat Psychoses. 2010;4(1):49-55.

11. Campana M, Falkai P, Siskind D, Hasan A, Wagner E. Characteristics and definitions of ultra-treatment-resistant schizo-phrenia - A systematic review and meta-analysis. Schizophr Res. 2021;228:218-26.

12. Polese D, Fornaro M, Palermo M, De Luca V, de Bartolomeis A. Treatment-Resistant to Antipsychotics: A Resistance to Everything? Psychotherapy in Treatment-Resistant Schizophrenia and Nonaffective Psychosis: A 25-Year Systematic Review and Exploratory Meta-Analysis. Front Psychiatry. 2019;10:210.

13. Thomas N, Hayward M, Peters E, van der Gaag M, Bentall RP, Jenner J, et al. Psychological Therapies for Auditory Hallu-cinations (Voices): Current Status and Key Directions for Future

Research. Schizophrenia Bulletin. 2014;40(Suppl_4):S202-S12.

14. Shergill SS, Murray RM, McGuire PK. Auditory hallucinations: a review of psychological treatments. Schizophr Res. 1998;32(3):137-50.

15. Pontillo M, De Crescenzo F, Vicari S, Pucciarini ML, Averna R, Santonastaso O, et al. Cognitive behavioural therapy for auditory hallucinations in schizophrenia: A review. World J Psychiatry. 2016;6(3):372-80.

16. Morrison AP, Pyle M, Gumley A, Schwannauer M, Turkington D, MacLennan G, et al. Cognitive behavioural therapy in clozapine-resistant schizophrenia (FOCUS): an assessor-blinded, randomised controlled trial. Lancet Psychiatry. 2018;5(8):633-43.

17. Shukla P, Padhi D, Sengar KS, Singh A, Chaudhury S. Efficacy and durability of cognitive behavior therapy in managing hallucination in patients with schizophrenia. Ind Psychiatry J. 2021;30(2):255-64.

18. Dellazizzo L, Potvin S, Phraxayavong K, Dumais A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. NPJ Schizophr. 2021;7(1):9.

19. Laws KR, Darlington N, Kondel TK, McKenna PJ, Jauhar S. Cognitive Behavioural Therapy for schizophrenia - outcomes for functioning, distress and quality of life: a meta-analysis. BMC Psychol. 2018;6(1):32.

20. Dellazizzo L, Giguère S, Léveillé N, Potvin S, Dumais A. A systematic review of relational-based therapies for the treat-ment of auditory hallucinations in patients with psychotic disorders. Psychol Med. 2022;52(11):2001-8.

21. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophr Res. 2018;197:176-81.

22. Aali G, Kariotis T, Shokraneh F. Avatar Therapy for people with schizophrenia or related disorders. Cochrane Database Syst Rev. 2020;5(5):Cd011898.

23. Beaudoin M, Potvin S, Machalani A, Dellazizzo L, Bourguignon L, Phraxayavong K, et al. The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clin Psychol Psychother. 2021;28(3):500-

18.

24. Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychoth-er. 2018;25(6):878-85.

25. Sebele-Mpofu FY, Serpa S. Saturation controversy in qualitative research: Complexities and underlying assumptions. A literature review. Cogent Social Sciences. 2020;6(1).

26. Habehh H, Gohel S. Machine Learning in Healthcare. Curr Genomics. 2021;22(4):291-300.

27. Usama M, Qadir J, Raza A, Arif H, Yau K-lA, Elkhatib Y, et al. Unsupervised Machine Learning for Networking: Tech-niques, Applications and Research Challenges. IEEE Access. 2019;7:65579-615.

28. Kung B, Chiang M, Perera G, Pritchard M, Stewart R. Unsupervised Machine Learning to Identify Depressive Subtypes. Healthc Inform Res. 2022;28(3):256-66.

29. Dwyer DB, Kalman JL, Budde M, Kambeitz J, Ruef A, Antonucci LA, et al. An Investigation of Psychosis Subgroups With Prognostic Validation and Exploration of Genetic Underpinnings: The PsyCourse Study. JAMA Psychiatry. 2020;77(5):523-33.

30. Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. Psy-chother Res. 2021;31(1):92-116.

31. Tanana MJ, Soma CS, Srikumar V, Atkins DC, Imel ZE. Development and Evaluation of ClientBot: Patient-Like Conversa-tional Agent to Train Basic Counseling Skills. J Med Internet Res. 2019;21(7):e12529.

32. Tang SX, Kriz R, Cho S, Park SJ, Harowitz J, Gur RE, et al. Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. NPJ Schizophr. 2021;7(1):25.

33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011;12:2825-30.

34. Hao J, Ho TK. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. Journal of Educational and Behavioral Statistics. 2019;44(3):348-61.

35. Camargo A. PCAtest: testing the statistical significance of Principal Component Analysis in R. PeerJ. 2022;10:e12967.

36. Yuan C, Yang H. Research on K-Value Selection Method of K-Means Clustering Algorithm. J. 2019;2(2):226-35.

37. Zubair M, Iqbal MDA, Shil A, Chowdhury MJM, Moni MA, Sarker IH. An Improved K-means Clustering Algorithm To-wards an Efficient Data-Driven Modeling. Annals of Data Science. 2022.

38. Zhou H, Zhang Y, Liu Y. A Global-Relationship Dissimilarity Measure for the k-Modes Clustering Algorithm. Comput Intell Neurosci. 2017;2017:3691316.

39. Wang M, Abrams ZB, Kornblau SM, Coombes KR. Thresher: determining the number of clusters while removing outliers. BMC Bioinformatics. 2018;19(1):9.

40. Yang X, Yang K, Cui T, Chen M, He L. A Study of Text Vectorization Method Combining Topic Model and Transfer Learn-ing. Processes. 2022;10(2).

41. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202.

42. O'Brien C, Rus-Calafell M, Craig TK, Garety P, Ward T, Lister R, et al. Relating behaviours and therapeutic actions during AVATAR therapy dialogue: An observational study. Br J Clin Psychol. 2021;60(4):443-62.

43. Fassbinder E, Schweiger U, Martius D, Brand-de Wilde O, Arntz A. Emotion Regulation in Schema Therapy and Dialectical Behavior Therapy. Front Psychol. 2016;7:1373.

44. Zarbo C, Tasca GA, Cattafi F, Compare A. Integrative Psychotherapy Works. Front Psychol. 2015;6:2021.

45. Cohen AS, Rodriguez Z, Warren KK, Cowan T, Masucci MD, Edvard Granrud O, et al. Natural Language Processing and Psychosis: On the Need for Comprehensive Psychometric Evaluation. Schizophr Bull. 2022;48(5):939-48.

46. Vaillant GE. Involuntary coping mechanisms: a psychodynamic perspective. Dialogues Clin Neurosci. 2011;13(3):366-70.

Figures and Tables

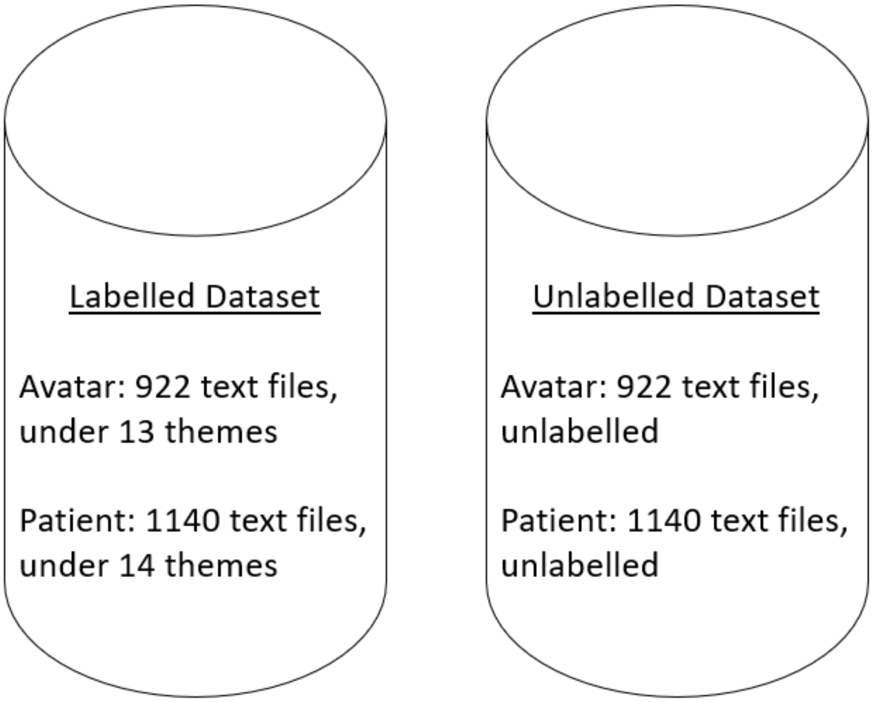
Labelled Dataset
Avatar: 922 text files, under 13 themes
Patient: 1140 text files, under 14 themes
Unlabelled Dataset
Avatar: 922 text files, unlabelled
Patient: 1140 text files, unlabelled

Figure 1. Datasets. The datasets contain text files of interactions between the Avatar and the Patient from the verbatims of immersive sessions. Each text file contains from 1 to 40 interactions. In the labelled dataset the text files are categorized as per one of Beaudoin et al. (2021) sub-themes, and in the unlabeled dataset, the text files are not categorized.

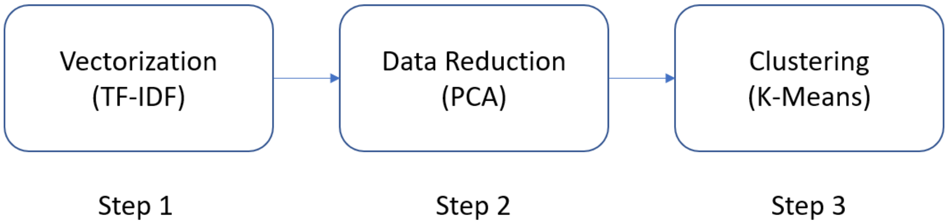

Figure 2. Overview of the steps performed to cluster the unlabeled data.

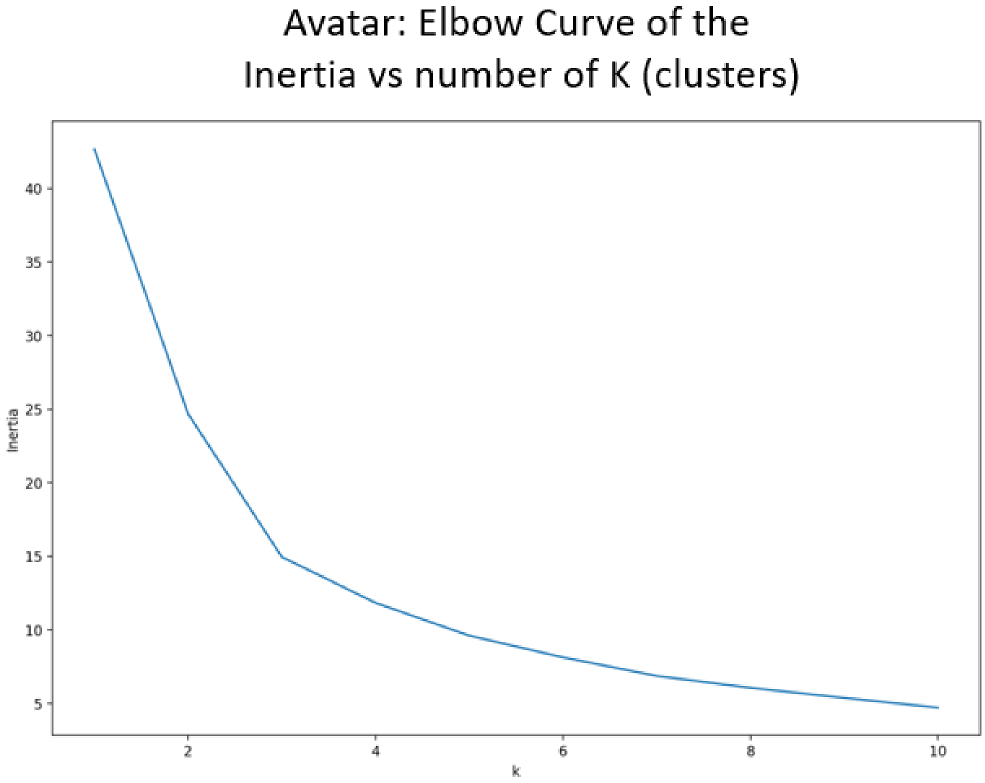

Figure 3. Elbow curve to identify the number of clusters for the Avatar interactions.

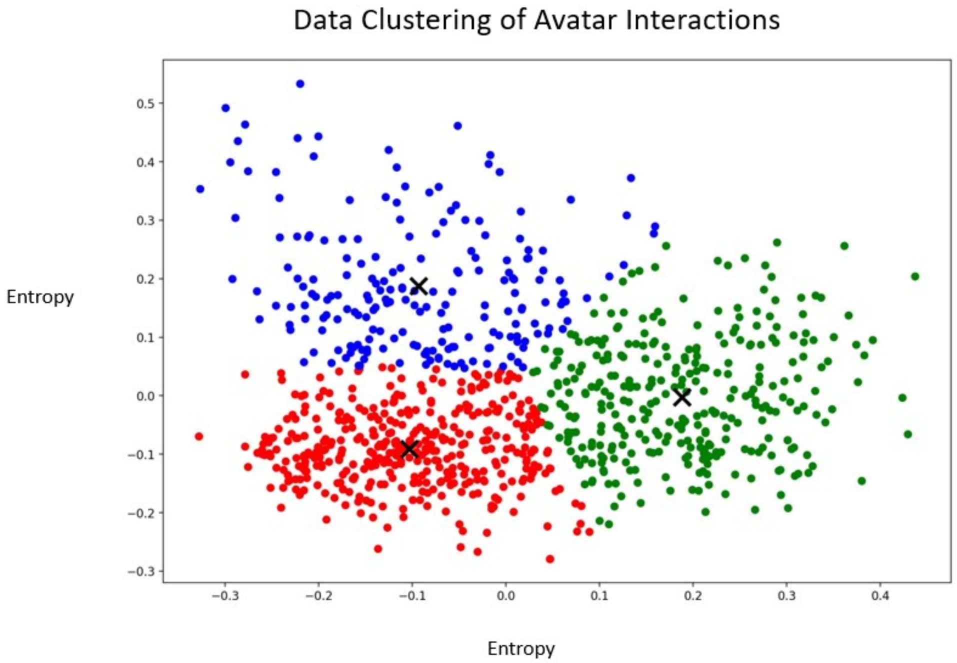

Figure 4. Graphical representation of the clusters containing the vectorized interactions of the Avatar. The X represent the centroids of the clusters.

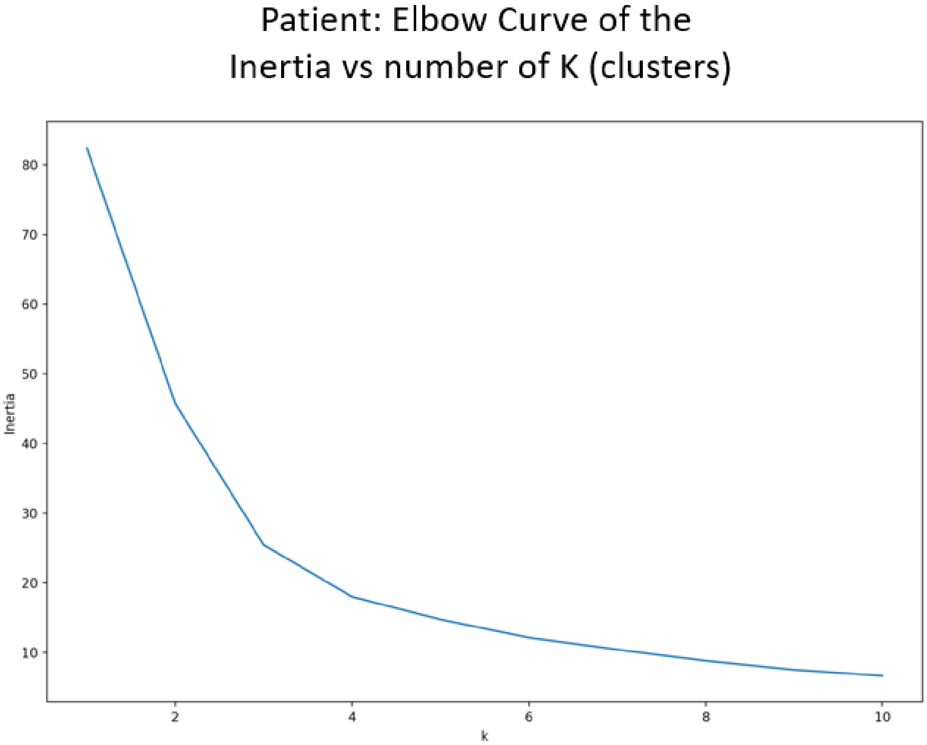

Figure 5. Elbow curve to identify the number of clusters for the Patient interactions.

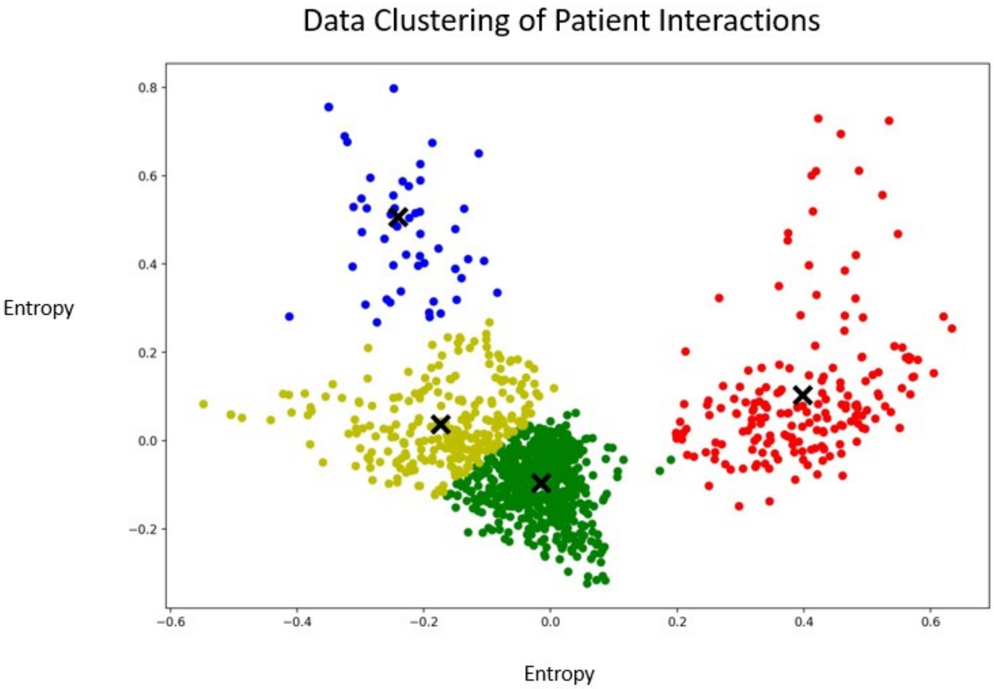

Figure 6. Graphical representation of the clusters containing the vectorized interactions of the

Patient. The X symbols represent the centroids of the clusters.

Table 1. Main themes (labels) and number of text files for the labelled dataset.

Main Themes (Avatar)	**Number of text files (n)**	**Percent of text files (%)**
Confrontational Techniques	427	46.31
Positive Techniques	495	53.69
Main Themes (Patient)	Number of text files (n)	Percent of text files (%)
Self-Perceptions	132	11.58
Aspirations	260	22.81
Emotional Responses	192	16.84
Coping Mechanisms	356	31.23
Beliefs About Voices and Schizophrenia	200	17.54

Table 2. Main themes (labels) and number of text files for the unlabeled dataset.

Cluster colors (Avatar)	**Mapping with the labelled data set**	**Number of text files (n)**	**Percent of text files (%)**
Blue	Confrontational Techniques	167	18.11
	Positive Techniques	72	7.81
Green	Confrontational Techniques	211	22.89
	Positive Techniques	127	13.77
Red	Confrontational Techniques	49	5.31
	Positive Techniques	296	32.10
Cluster colors (Patient)	Mapping with the labeled data set	Number of text files (n)	Percent of text files (%)
Blue	Self-Perceptions	12	1.05
	Aspirations	8	0.70
	Emotional Responses	6	0.53
	Coping Mechanisms	2	0.18
	Beliefs About Voices and Schizophrenia	10	0.88
Yellow	Self-Perceptions	83	7.28
	Aspirations	95	8.33
	Emotional Responses	109	9.56
	Coping Mechanisms	79	6.93
	Beliefs About Voices and Schizophrenia	43	3.77
Green	Self-Perceptions	33	2.89
	Aspirations	119	10.44
	Emotional Responses	67	5.88
	Coping Mechanisms	237	20.79

	Beliefs About Voices and Schizophrenia	144	12.63
Red	Self-Perceptions	4	0.35
	Aspirations	38	3.33
	Emotional Responses	10	0.88
	Coping Mechanisms	38	3.33
	Beliefs About Voices and Schizophrenia	3	0.26

Article 7. Ensemble Methods to Optimize Automated Text Classification in Avatar Therapy

Alexandre Hudon

Kingsada Phraxayavong

Stéphane Potvin

Alexandre Dumais

Accepté après corrections mineures dans *BioMedInformatics (2024).* Article révisé par les pairs.

Abstract

Background: Psychotherapeutic approaches such as Avatar Therapy (AT) are novel therapeutic attempts to improve patients diagnosed with treatment-resistant schizophrenia. Qualitative analyses of immersive sessions of AT has been undertaken to enhance and refine the existing interventions taking place in this therapy. To account for the time-consuming, costly, and potential misclassification biases, prior implementation of a Linear Support Vector Classifier provided helpful insight. Single model implementation for text classification is often limited, especially for dataset containing imbalanced data. The main objective of this study is to evaluate the change in accuracy of automated text classification machine learning algorithms when using an ensemble approach on immersive session verbatims of AT. Methods: An ensemble model, comprising five machine learning algorithms was implemented to conduct text classification for avatar and patient interactions. The models included in this study are: Multinomial Naïve Bayes, Linear Support Vector Classifier, Multi-layer perceptron classifier, XGBClassifier and the K-Nearest-Neighbor model. Accuracy, precision, recall and f1-score were compared for the individual classifiers and the ensemble model. Results: The ensemble model performed better than its individual counterparts for accuracy. Conclusion: Using an ensemble methodological approach, this methodology might be employed in future research to provide insight into the interactions being categorized and the therapeutical outcome of patients based on their experience with AT with optimal precision.

Keywords

Virtual Reality Therapy; Artificial intelligence; Auditory Hallucinations; Schizophrenia, Psychotherapy; Machine learning; Ensemble modeling; Text classification

Introduction

Schizophrenia is a complex psychopathology characterized by positive symptoms (such as auditory hallucinations, persecutory delusions), disorganization of thoughts and behaviors, and negative symptoms (i.e.: such as avolition, anhedonia, alogia) [1, 2]. Pharmacological treatment of schizophrenia focuses primarily on positive symptoms because they can be linked to serious deleterious events (such as suicide and violence) [3, 4]. However, recent studies have demonstrated that around 25 to 30% of patient are resistant to current lines of treatment [5-7]. Multiple definition exists when referring to treatment resistant [8]. The most common accepted definition is when, after two trials with antipsychotic medicines with verified adherence and an appropriate dose and duration, symptoms persisted [9]. Patients who meet this definition are known as patients with treatment resistant schizophrenia (TRS). Clozapine, a second-generation antipsychotic, is commonly used to treat patients with TRS [10]. Up to 60 percent of patients on clozapine respond poorly to this approach, which is why further approaches, notably adjunct to medication, have been used or are currently being studied [11].

Once such approach is psychotherapy. Amongst psychotherapeutic approaches, cognitive behavioral therapy (CBT) is one of the most used [12]. Despite statistical improvements of patients, little evidence has been found to recommend its use routinely in patients suffering from TRS considering the lack of evidence in clinical improvements [13]. Therefore, further therapeutical approaches are currently being studied, such as Avatar Therapy (AT) [14]. The effectiveness of this treatment in lowering patients' resistant auditory hallucinations and gauging their wellbeing is still being investigated [15]. In order to interact with patients in an immersive setting, AT recommends utilizing a virtual reality headset. [16]. In AT, the therapist simulates the patient's auditory hallucination while using the virtual environment by the mean of an animated visual depiction (pre-configured in collaboration with the patient). The Leff et al. (2014) team in London, United Kingdom, created AT in 2008 [17]. The first randomized controlled trial, single-blinded, was conducted in South London and Maudsley NHS Trust (United Kingdom) in 2016 with 150 adult patients who had been clinically diagnosed with schizophrenia spectrum but continued to experience auditory verbal hallucinations despite receiving treatment [18]. AT or supportive therapy was randomly assigned to these patients. Evaluated by the Psychotic Symptoms Rating Scales Auditory Hallucinations (PSYRATS-AH), the primary result was a decrease in auditory verbal hallucinations, as well as a decrease in depressive symptoms at 12 weeks [18]. A current clinical trial at the University Institute in Mental Health of Montreal. (IUSMM) that compares CBT to AT for patients with schizophrenia who are receiving ongoing therapy and experiencing auditory hallucinations is currently taking place. In this study comprising 136 patients, 68 are receiving AT and 68 are receiving CBT. While this experiment is being conducted, 37 patients who participated in AT and 37 who participated in CBT were evaluated during the course of a 365 days pilot randomized comparison trial at the IUSMM to determine the efficacy of AT over CBT for this population. For these individuals, AT performed better than CBT did on auditory hallucinations, and it also significantly improved quality of life and persecutory beliefs [19].

To assess the content of AT and provide a more comprehensive grasp of the dynamics taking place betwixt the patient and their avatar during the immersive sessions, qualitative analyses have been conducted. A preliminary content analysis of AT was performed in 2018 by Dellazizzo et al., who explored the treatment of 12 AT patients. A total of 5 themes emerged from patients' conversations with the avatar, according to this analysis: emotional response to voices, ideas about voices and schizophrenia, oneself, coping techniques, and goals [20]. This analysis provided a first insight about prospective AT treatment targets. In a follow-up study done in 2021, Beaudoin et al. qualitatively evaluated 125 therapy verbatims of patients who received AT. The avatar's two main key interactions' themes were confrontational techniques (which had eight sub-themes) and positive techniques (which included six sub-themes). The patients' emotional reactions, self-perceptions, coping strategies, goals and notions of voices and schizophrenia were all highlighted as five key themes. A total of 14 sub-themes were identified amongst these five main themes [21]. By illuminating the interactions between avatars and patients, it was possible to highlight important areas of focus that may direct future research and therapeutic interventions, these descriptive investigations advance our understanding of the therapeutic process in AT. While descriptive data may offer extensive insights, it lacks the quantitative counterpart necessary for identifying the precise components of psychotherapy that may help patients achieve favorable outcomes [22]. Qualitative analyses are also costly and time-consuming and are subject to inherent biases such as misclassification bias, which is even more prevalent when different kinds of interaction can overlap (which is frequent in natural language) [23-25].

To provide a quantitative propensity to qualitative analyses of such verbatims, classification techniques can be used [26]. This is usually done using machine learning algorithms. Classification techniques can be supervised (data is deduced from a labeled dataset) or unsupervised (data is inferred) [27]. One major problem is often that such implementations need large datasets to derive accurate classification [28]. Another limitation is the limited data availability in the psychotherapeutic setting considering the confidential nature of the interactions between patients and their therapists. Recent literature review on machine learning algorithms on small datasets in the context of psychotherapy identified several key algorithms that can perform acceptably on such datasets [29]. A first implementation of such technique on AT verbatims was done using a linear vector classifier (LSVC) as per this literature review and concluded that it can conduct automated theme classifications on AT session transcripts using a limited dataset, achieving accuracy and substantial classification agreement comparable to that of human coders. [30]. This technique was also found useful to efficiently identify interactions between the avatar and the patient in AT [31]. However, this approach is limited by the linear assumptions of LSVC (i.e: interactions in AT are assumed to be entirely linearly separable), by its sensitivity to data outliers and by the difficulty in successfully classify imbalanced data [32]. The AT dataset contains imbalanced data considering that some type of interactions between the patient and the avatar are more frequent.

To account for single model limitations, ensemble modeling is a technique that is widely used [33]. It consists in creating a better, more precise, and more reliable predictive model by combining the predictions of various distinct models [34]. This is done to in-crease the overall effectiveness and generalization of the ensemble by taking use of the diversity of predictions made by the individual models [35]. Such an approach can increase the model complexity and computational resources needed to be performed [33]. However, considering the small datasets employed for automatic classification in AT, these limitations are insignificant compared to the potential expected. To our knowledge, this has never been conducted on psychotherapeutic content and was never attempted on AT verbatims.

The main objective of this study is to assess the change in accuracy of automated text classification machine learning algorithms when using an ensemble approach on immersive session verbatims of AT. It is hypothesized that such an approach will increase the accuracy in automated text classification in AT and is therefore going to yield better automated text classification for future analyses of verbatims for patients who are receiving AT. When taking into account large amount of variables being incorporated into the automated classification of the interactions occurring in the verbatims for AT, a combination of different machine learning classification models is believed to account for potential misclassification as compared to the use of a single classification model.

Materials and Methods

Recruitment and participants

The dataset utilized in this investigation consists of therapeutic interactions of participants involved in a pilot trial carried out at the Research Center of the University Institute in Mental Health of Montreal (CR-IUSMM). They were all affected by treatment-resistant schizophrenia (TRS), which is marked by a continuous auditory hallucination despite the use of two or more dopaminergic antagonists. Their AT sessions were conducted between 2017 and 2022. The clinical trial can be found on ClinicalTrials.gov under the identifier NCT03585127. [19]. Each participant underwent a series of nine one-hour psychotherapeutic sessions, with eight of them being immersive sessions that included interaction with a virtual representation of their auditory verbal hallucinations known as the Avatar. The study comprised individuals who were over 18 years old and were patients at the IUSMM.

Dataset: Corpus of Avatar Therapy and features

Research assistants transcribed verbatim the immersive sessions of 18 AT patients from audio recordings. Subsequently, AH reviewed the verbatims to ensure transcription accuracy, yielding a total of 144 verbatims, representing nearly 70 hours of AT immersion. Interactions between patients and avatars were annotated and categorized based on the 27 themes specified in Beaudoin et al. (2021). Table 1 presents the categorized interaction themes for both the avatar and the patients.

-- Please insert Table 1 here--

A dataset was created using 144 therapy transcripts from 18 randomly chosen patients who received AT at the CR-IUSMM between 2017 and 2022. Eight treatment sessions were attended by each patient, resulting in an average of eight transcripts per subject. The initial transcripts were meticulously typed in Canadian French. To annotate the transcripts, the twenty-seven themes listed in the study done by Beaudoin et al. (2021) were applied manually [21]. The software QDA Miner version 5, developed by Provalis Research, was utilized for qualitative data analysis for the annotation process [36]. The annotations were then retrieved into text files, each of which contained one to forty inter-actions that were all related to the same theme. According to the model depicted in Figure 1, the annotations retrieved were subsequently separated into two conceptual databases: one for the avatar and one for the patient. The different themes found in the dataset were balanced.

-- Please insert Figure 1 here--

Ensemble modelling

Ensemble modelling implies the use of several classification models to select the best performing model according to a vote for each classification conducted. In this study, the models being implemented as part of the ensemble are as follows: LSVC, Multinomial Naïve Bayes (Multinomial NB), Multi-layer perceptron classifier (MLP), XGBClassifier (XGB) and the K-Nearest-Neighbor (KNN) algorithms. These were selected based on the previous literature review on best performing algorithms on small datasets and the ground rules for composing an ensemble model, notably using diverse base models, potential for cross-validation and avoidance of highly correlated models. The ensemble model is presented in Figure 2. Ensemble modeling functions were selected and used from the Scikit-Learn library with Python 3.11 [37].

-- Please insert Figure 2 here --

Both the patient conceptual dataset and the avatar conceptual dataset were employed by the ensemble model. To refine classification techniques and optimize the machine learning algorithm's performance, a GridSearchCV (GSCV) approach from the Scikit-Learn library was implemented [37]. Users can explore various hyperparameters and cross-validate the classifier's predictions using the GSCV tool to identify the optimal set of parameters that yield the highest performance. In this study, LSVC classifiers were both subjected to GSCV. The MLP, Multinomial NB classifiers and XGB performed better when considering hyperparameterization, hence default values were used for these. The KNN was initialized with the default Minkowski distance of three, which is consequent with previous instantiation and analysis of KNN performances on AT [38].

The term frequency-inverse document frequency (TF-IDF) method, which outperforms other algorithm-tokenizer combinations in text categorization, was used in conjunction with the algorithms [39]. The implementation of TfidfVectorizer, offered by the Scikit-Learn module, to implement TF-IDF tokenization, was used [37]. The raw text retrieved from the therapeutic interactions between the avatar and the patient during immersive session were converted into numerical vectors.

Machine learning algorithms

The five models used to compose the ensemble model are listed below.

1. LSVC: The Support Vector Machine (SVM) approach aims to determine the optimal hyperplane for dividing various classes of data points in a high-dimensional feature space. This involves maximizing the margin between classes to achieve robust generalization performance. The method identifies support vectors, a subset of training samples serving as pivotal points for

the decision boundary. Unlike SVC, the LSVC uses a linear kernel. A kernel is a mathematical function transforming data into a higher-dimensional feature space, crucial in handling complex problems that may be challenging or impossible in the original input space. A linear kernel is applied when data separation can be achieved linearly. The implementation of SVC in this study utilized Scikit-Learn, specifically the SVC class from the SVM library [37].

2. The Multinomial Naive Bayes method is a derivative of the Naïve Bayes technique, which, based on the Bayes theorem, assumes conditional independence of features given the class. This method is developed using the Bayes theorem, which enables the updating of the probability of Event A occurring in light of new information or additional supporting evidence from Event B. It calculates the posterior probability (P(A|B)) by combining the prior probability (P(A)) and the likelihood (P(B|A)). Specifically designed to handle discrete features in text data, such as word counts or frequencies, the Multinomial Naive Bayes classifier is implemented using Scikit-Learn, with the MultinomialNB class from the naive_bayes module being employed in this study [37].

3. MLP: The Multi-Layer Perceptron (MLP) classifier serves for classification and various machine learning applications. It constitutes a feedforward neural network model characterized by multiple layers of interconnected neurons. The typical structure of the MLP classifier includes an input layer, one or more hidden layers, and an output layer. Each layer comprises numerous neurons that conduct computations on incoming data and transmit the results to the subsequent layer. In an MLP, every neuron in each layer is connected to every other neuron in both the layers above and below, indicating full connectivity. The strengths and relevance of information flow across the network are influenced by weights associated with neuron connections. In this study, the MLP implementation is derived from Scikit-Learn, specifically utilizing the MLPClassifier class from the neural_network library [37].

4. XGB: XGB works by sequentially iteratively generating an ensemble of weak learners such as decision trees. Each successive model is then used to correct the mistakes of the prior ones. This is done to optimize the ensemble's predictive capability while reducing overfitting by

minimizing a user-defined loss function. XGboost library, namely the XGBoost class of the XGBClassifier module, provided the XGB implementation used in this study [37].

5. KNN: This widely employed technique is based on the principles of k-means clustering, aiming to group similar data in a comprehensible, relatively swift, and scalable manner while ensuring convergence. It assesses whether two items are identical and organizes them based on their Euclidean distance, representing the length of a line drawn between two data points. The number of clusters is predetermined, and the process un-folds iteratively. Beginning with the random selection of the center (centroid) for each cluster, the Euclidean distance from all data points to the centroids is computed, and the data points are assigned to their closest clusters. Subsequently, for each cluster, a new centroid is determined by calculating the mean of all data points within the cluster. This process is repeated until all points converge, and the cluster centers cease to move. The KNN implementation in this study is sourced from Scikit-Learn, specifically utilizing the neighbors class from the SVM library [37].

Voting technique

Ensemble modeling offers different avenues to compare predictions of the classification models they encompass. This technique is useful as classification depends on the performance of multiple models and will therefore not be hampered by big errors or misclassifications from a single model. A bad performance from one model can be compensated for by a great performance from another. One of the most widely used technique to assess the accuracy of ensemble models is the voting process [40].

There exist two categories of voting techniques: hard and soft voting [41]. In voting techniques applied to AT, hard voting implies adding up all the forecasts for each interaction themes and predicting the interaction theme with the most votes [42]. Soft voting in-volves summing the anticipated probabilities (accuracy scores) for each interaction theme and predicting the theme with the highest probability [43]. Hard voting is useful when the models in the ensemble are diverse and do not give well-calibrated probabilities, as com-pared to soft voting (which is better at capturing the nuances of different models' confidence levels) [41]. Considering that nuances of the different models' confidence levels is needed as data cannot be assumed to be well balanced across the different interaction themes, soft voting is used in this study.

Data analysis and validation

Data regarding the classification performance of each theme, encompassing recall, accuracy, and f1-Score for each algorithm, was compiled using the Classification Report tool within the Scikit-Learn metrics module [37]. The overall average accuracy is established by the ratio of true predictions to total predictions. The f1-Score assesses the accuracy of theme classification, with recall indicating the sensitivity of the prediction, and precision reflecting the positive predictive value for each prediction [44]. To offer a com-prehensive evaluation of classification accuracy, the f1-Score—a commonly used measure in text classification—strikes a balance between precision and recall. Hence, the f1-Score is the harmonic mean of recall and precision [45].

For each conceptual database, a partitioning approach was employed, allocating 70% of the annotated texts for training the algorithms and reserving the remaining 30% for testing purposes [46]. The objective was to establish a statistical likelihood for each algorithm, expressed through a prediction score, indicating the efficacy of classifying interactions. To adhere to recommended design practices, the training and testing sets were deliberately kept distinct for the calibration of each machine learning classifier [47]. Additionally, a tenfold cross-validation strategy was implemented for each algorithm, utilizing the K-Fold model from the Scikit-Learn module [37].

Results

Characteristics of the participants

The interactions occurring in the verbatims of 35 patients were utilized by the five machine learning algorithms in this study for automated annotation. The details of the sampled patients can be found in Table 2.

--Please insert Table 2 here --

Performance of Ensemble Modeling

The performance in accuracy of individual models and ensemble model for the automated classification of avatar' interactions and patient' interactions are found below.

Automated classification of avatar interactions

The accuracy scores are presented in Figure 3 for all the individual models on top of the ensemble model. The ensemble model performed the best with cross-validated accuracy of 0.71, closely followed by the LSVC at 0.66 and the MLP classifier at 0.66. The XGB performed with a cross-validated accuracy of 0.54 and the KNN algorithm performed with an accuracy of 0.57. The Multinomial NB performed the worst with an accuracy of 0.48.

-- Please insert Figure 3 here--

Mean metrics for recall, precision and f1-score for classification of avatar interactions are presented in Table 3. It can be observed that the performances of all the metrics is consistent with the findings explicated for the accuracy, with the ensemble model achieving the best performance for accuracy, recall , precision, and f1-score. This is closely followed by the LSVC and the MLP classifiers.

-- Please insert Table 3 here--

Automated classification of patient interactions

The accuracy scores are presented in Figure 4 for all the individual models coupled with the ensemble model. The ensemble model performed the best with cross-validated accuracy of 0.58. This is almost tied with the LSVC at 0.57 and the MLP classifier at 0.54. The XGB performed with a cross-validated accuracy of 0.48 and the KNN algorithm per-formed with an accuracy of 0.45. The Multinomial NB performed the worst with an accuracy of 0.44.

-- Please insert Figure 4 here--

Mean metrics for recall, precision, and f1-score for classification of patient interactions are presented in Table 4. It can be observed that the performances of almost all the metrics is consistent with the findings explicated for the accuracy, with the ensemble model achieving the best performance for accuracy, recall, precision and f1-score. This is closely followed by the LSVC and the MLP classifiers. The only divergent metric found is that the LSVC performs better than the ensemble model for precision.

-- Please insert Table 4 here--

Discussion

The main objective of this study was to evaluate the change in accuracy of automated text classification machine learning algorithms when using an ensemble approach on immersive session verbatims of AT. For automatic classification of avatar and patient interactions, the ensemble approach performed best in terms of the classification accuracy. This was also the case for the recall, precision and f1-score metrics, apart from precision (in the classification of patient interactions) which was found to be better with the LSVC.

The performance of ensemble modeling approach is often preferred as compared to a single model when the data used is complex. This can hardly be compared to literature on the context of psychotherapy considering this has not been performed previously. How-ever, as an example in the context of text classification, a recent study comparing the performances of several machine learning algorithms to an ensemble model comprising these algorithms, on a corpus comprising of the Youtube Spam Collection Dataset and different text vectorization approaches, demonstrated that some of the ensemble (such as Adaboost and LightGBM) learning methods frequently produce enhanced text classification performance compared with base techniques [48]. It is also to be noted that literature reviews on ensemble methods highlights that ensemble modeling is an acceptable technique for coping with individual classifiers' large variation while minimizing general mistakes [49]. Furthermore, ensemble techniques are reported to be an appropriate method to improve accuracy in text classification tasks which is what has been observed in its use of AT [50]. Interestingly, a recent study comparing the use of single classifier to an ensemble approach in the domain of mental health suggests that for the prediction of mental health problems, ensemble models demonstrate better prediction results [51]. This could be similar for the appropriate prediction of patient interactions in the setting of psychotherapy as this is likewise established on textual instances.

The performance of LSVC for avatar and patient interactions was very close as those of the ensemble model. LSVC was also found to perform better for precision for the patient interactions. This could be explained by the fact that most of the data was in fact linearly separable as it was previously assumed. Considering this sort of separability of the data, the data diversity is decreased and therefore the ensemble model compares to the best performing linear classifier model, which is in this scenario the LSVC [52]. It could be hypothesized that the small amount of patients' data presented in the dataset also accounts for this observation considering that, as more data becomes available, new themes could emerge from the verbatims and account for multicollinearity. This can be seen if two or more variables have linear correlations, which entails that determining the marginal in-fluence of a variable will be difficult [53].

The classification performances of the algorithms on the avatar conceptual dataset compared to the patient dataset indicated that interactions involving the avatar were classified with a higher overall accuracy. This can be explained by the fact that the classification complexity is reduced for the avatar as there are 13 possible themes for the classification as compared to 14 for the patient interactions.

Potential future applications of ensemble modeling to the field of psychotherapy could achieve similar results as other ensemble modeling techniques in clinical psychiatry. For example, machine learning applications of ensemble models for clinical information of 685 outpatients enabled the prediction of successful outcome of cognitive behavioral therapy with a balanced accuracy of 69% [54]. This sort of accuracy is comparable to the ones observed in our study. However, considering the limited number of studies that applies machine learning to psychotherapeutic content, it is clear at this stage that future studies are needed, notably on textual entities such as therapeutic interactions.

It is also important to note practical ethical considerations when using such techniques for psychotherapeutic interventions. In this study, considering that data was pro-cessed by several machine learning algorithms, data was anonymized to ensure confidentiality and privacy of the patients. The accountability of data being automatically categorized is also the responsibility of the clinician when machine learning is applied to a clinical context and this should be further investigated [55].

Limitations

The models utilized in constructing the ensemble model are currently limited by the relatively small databases available for Avatar Therapy (AT). The performance trend of the ensemble model will be re-evaluated as more patients are added to the dataset. It's important to note that the transcripts analyzed in this study were written in Canadian French, and obtaining vectorizers that included stop-words specific to Canadian French proved challenging. Stop-words, which are often excluded during tokenization due to their limited meaning, may impact the analysis's accuracy. The lack of sufficient stop-words for Canadian French could result in the inclusion of inconsequential terms. Regarding the patient conceptual database, it is noteworthy that three-fifths of the individual algorithms initially achieved a classification accuracy below 0.5, limiting the performance of the ensemble models.

Conclusion

In conclusion, this study evaluated the change in accuracy of automated text classification machine learning algorithms when using an ensemble approach on immersive session verbatims of AT. Automated classification of textual is not a simple task when considering psychotherapeutic interventions and this study demonstrated that ensemble modeling performed best in terms of accuracy for classification of avatar and patient interactions. This technique performed also better than its individual counterparts for precision, recall and f1-score. The only exception being the precision in the classification of patient interactions, for which the LSVC performed best. This study offers a first evaluation of ensemble modeling in the context of AT and provided an objective optimized approach in the classification of textual interactions based on immersive session verbatims. This technique might be used in future research to give insight into the interactions being classified and the therapeutical response of patients based on their experience with AT immersion sessions with optimized precision by employing an ensemble methodological approach.

Author Contributions

Conceptualization, A.H., K.P., S.P. and A.D.; methodology, A.H. and A.D.; validation, A.H. and A.D.; formal analysis, A.H.; investigation, A.H.; data curation, A.H.; writing—original draft preparation, A.H.; writing—review and editing, A.H., K.P, S.P. and A.D.; supervision, K.P., S.P. and A.D.; project administration, K.P.; funding acquisition, K.P., S.P. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work was indirectly supported by Le Fonds de recherche du Québec – Santé (FRQS); Otsuka Canada Pharmaceutical Inc.; Chaire Eli Lilly Canada de recherche en schizophrénie; MEI (Ministère de l’Économie et de l’Innovation); Services et recherches psychiatriques AD; Fonds d’excellence en recherche Apogée Canada. These fundings bodies had no part in the data collection, analysis, interpretation of data and in writing the manuscript.

Institutional Review Board Statement

This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients. Patients that are part of this study were selected based on the proof-of-concept trial from Percy du Sert 2018’s study and Dellazizzo 2021’s study [15]. The trial was conducted in accordance with the Declaration of Helsinki and was approved by the institutional ethical committee (CER IPPM 16-17-06). We obtained written informed consent from all patients.

Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available due to patients’ privacy but are available from the corresponding author on reasonable request.

Acknowledgments: "Database icon" by Nurul Hotimah is licensed under Creative Commons BY 3.0 (https://creativecommons.org/licenses/by/3.0/).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Carrà G, Crocamo C, Angermeyer M, Brugha T, Toumi M, Bebbington P. Positive and negative symptoms in schizophrenia: A longitudinal analysis using latent variable structural equation modelling. Schizophr Res. 2019;204:58-64.

2. Correll CU, Schooler NR. Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. Neuropsychiatr Dis Treat. 2020;16:519-34.

3. Patel KR, Cherian J, Gohil K, Atkinson D. Schizophrenia: overview and treatment options. P t. 2014;39(9):638-45.

4. Stępnicki P, Kondej M, Kaczor AA. Current Concepts and Treatments of Schizophrenia.

Molecules. 2018;23(8):2087.

5. Siskind D, Siskind V, Kisely S. Clozapine Response Rates among People with Treatment-Resistant Schizophrenia: Data from a Systematic Review and Meta-Analysis. Can J Psychiatry. 2017;62(11):772-7.

6. Corripio I, Roldán A, Sarró S, McKenna PJ, Alonso-Solís A, Rabella M, et al. Deep brain stimulation in treatment resistant schizophrenia: A pilot randomized cross-over clinical trial. EBioMedicine. 2020;51:102568.

7. Huckle PL, Palia SS. Managing resistant schizophrenia. Br J Hosp Med. 1993;50(8):467-71.

8. Suzuki T, Remington G, Mulsant BH, Uchida H, Rajji TK, Graff-Guerrero A, et al. Defining treatment-resistant schizophrenia and response to antipsychotics: a review and recommendation. Psychiatry Res. 2012;197(1-2):1-6.

9. Correll CU, Howes OD. Treatment-Resistant Schizophrenia: Definition, Predictors, and Therapy Options. J Clin Psychiatry. 2021;82(5).

10. Chakrabarti S. Clozapine resistant schizophrenia: Newer avenues of management. World J Psychiatry. 2021;11(8):429-48.

11. Potkin SG, Kane JM, Correll CU, Lindenmayer J-P, Agid O, Marder SR, et al. The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. npj Schizophrenia. 2020;6(1):1.

12. Bighelli I, Huhn M, Schneider-Thoma J, Krause M, Reitmeir C, Wallis S, et al. Response rates in patients with schizophrenia and positive symptoms receiving cognitive behavioural therapy: a systematic review and single-group meta-analysis. BMC Psychiatry. 2018;18(1):380.

13. Morrison AP, Pyle M, Gumley A, Schwannauer M, Turkington D, MacLennan G, et al. Cognitive behavioural therapy in clozapine-resistant schizophrenia (FOCUS): an assessor-blinded, randomised controlled trial. Lancet Psychiatry. 2018;5(8):633-43.

14. Aali G, Kariotis T, Shokraneh F. Avatar Therapy for people with schizophrenia or related disorders. Cochrane Database Syst Rev. 2020;5(5):Cd011898.

15. Leff J, Williams G, Huckvale MA, Arbuthnot M, Leff AP. Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. Br J Psychiatry. 2013;202:428-33.

16. Bisso E, Signorelli MS, Milazzo M, Maglia M, Polosa R, Aguglia E, et al. Immersive Virtual Reality Applications in Schiz-ophrenia Spectrum Therapy: A Systematic Review. Int J Environ Res Public Health. 2020;17(17).

17. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff AP. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? Psychosis. 2014;6(2):166-76.

18. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry. 2018;5(1):31-40.

19. Dellazizzo L, Potvin S, Phraxayavong K, Dumais A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. NPJ Schizophr. 2021;7(1):9.

20. Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychother. 2018;25(6):878-85.

21. Beaudoin M, Potvin S, Machalani A, Dellazizzo L, Bourguignon L, Phraxayavong K, et al. The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clin Psychol Psychother. 2021;28(3):500-18.

22. Szymańska A, Dobrenko K, Grzesiuk L. Characteristics and experience of the patient in psychotherapy and the psycho-therapy's effectiveness. A structural approach. Psychiatr Pol. 2017;51(4):619-31.

23. Runciman WB. Qualitative versus quantitative research — balancing cost, yield and feasibility. Quality and Safety in Health Care. 2002;11(2):146-7.

24. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. Plast Reconstr Surg. 2010;126(2):619-25.

25. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211-7.

26. Dogra V, Verma S, Kavita, Chatterjee P, Shafi J, Choi J, et al. A Complete Process of Text Classification System Using State-of-the-Art NLP Models. Comput Intell Neurosci.

2022;2022:1883698.

27. Jovel J, Greiner R. An Introduction to Machine Learning Approaches for Biomedical Research. Front Med (Lausanne). 2021;8:771607.

28. Hey T, Butler K, Jackson S, Thiyagalingam J. Machine learning and big scientific data. Philos Trans A Math Phys Eng Sci. 2020;378(2166):20190054.

29. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms. JMIR Ment Health. 2021;8(10):e22651.

30. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach. Health Informatics Journal. 2022;28(4):14604582221142442.

31. Hudon A, Couture J, Dellazizzo L, Beaudoin M, Phraxayavong K, Potvin S, et al. Dyadic Interactions of Treatment-Resistant Schizophrenia Patients Having Followed Virtual Reality Therapy: A Content Analysis. Journal of Clinical Medicine. 2023;12(6):2299.

32. Bhavsar H, Ganatra A. A comparative study of training algorithms for supervised machine learning. International Journal of Soft Computing and Engineering (IJSCE). 2012;2(4):2231-307.

33. Ganaie MA, Minghui H, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence. 2022;115:105151.

34. Verma A, Mehta S, editors. A comparative study of ensemble learning methods for classification in bioinformatics. 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence; 2017: IEEE.

35. Pintelas P, Livieris IE. Special Issue on Ensemble Learning and Applications. Algorithms. 2020;13(6):140.

36. Lewis RB, Maas SM. QDA Miner 2.0: Mixed-model qualitative data analysis software. Field methods. 2007;19(1):87-108.

37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011;12:2825-30.

38. Hudon A, Beaudoin M, Phraxayavong K, Potvin S, Dumais A. Unsupervised Machine Learning Driven Analysis of Verbatims of Treatment-Resistant Schizophrenia Patients Having

Followed Avatar Therapy. J Pers Med. 2023;13(5).

39. Chen J, Yuan P, Zhou X, Tang X, editors. Performance Comparison of TF*IDF, LDA and Paragraph Vector for Document Classification2016; Singapore: Springer Singapore.

40. Kabari LG, Onwuka UC. Comparison of bagging and voting ensemble machine learning algorithm as a classifier. Interna-tional Journals of Advanced Research in Computer Science and Software Engineering. 2019;9(3):19-23.

41. Peppes N, Daskalakis E, Alexakis T, Adamopoulou E, Demestichas K. Performance of machine learning-based multi-model voting ensemble methods for network threat detection in agriculture 4.0. Sensors. 2021;21(22):7475.

42. Alsulami B, Almalawi A, Fahad A. Toward an Efficient Automatic Self-Augmentation Labeling Tool for Intrusion Detection Based on a Semi-Supervised Approach. Applied Sciences. 2022;12(14):7189.

43. Manconi A, Armano G, Gnocchi M, Milanesi L. A Soft-Voting Ensemble Classifier for Detecting Patients Affected by COVID-19. Applied Sciences. 2022;12(15):7554.

44. Goutte C, Gaussier E, editors. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. European conference on information retrieval; 2005: Springer.

45. Opitz J, Burst S. Macro f1 and macro f1. arXiv preprint arXiv:191103347. 2019.

46. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. 2018.

47. Birba DE. A Comparative study of data splitting algorithms for machine learning model selection. 2020.

48. Ibrahim Y, Okafor E, Yahaya B, Yusuf SM, Abubakar ZM, Bagaye UY. Comparative Study of Ensemble Learning Techniques for Text Classification. 2021. p. 1-5.

49. Ammar M, Rania K. An effective ensemble deep learning framework for text classification. Journal of King Saud University - Computer and Information Sciences. 2022;34(10, Part A):8825-37.

50. Palanivinayagam A, El-Bayeh CZ, Damaševičius R. Twenty Years of Machine-Learning-Based Text Classification: A Sys-tematic Review. Algorithms. 2023;16(5):236.

51. Chung J, Teo J. Single classifier vs. ensemble machine learning approaches for mental

health prediction. Brain Informatics. 2023;10(1):1.

52. Dietterich TG, editor Ensemble methods in machine learning. International workshop on multiple classifier systems; 2000: Springer.

53. Chan JY-L, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong Z-W, et al. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. Mathematics. 2022;10(8):1283.

54. Taubitz FS, Büdenbender B, Alpers GW. What the future holds: Machine learning to predict success in psychotherapy. Behav Res Ther. 2022;156:104116. doi:10.1016/j.brat.2022.104116

55. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bull World Health Organ. 2020;98(4):251-256. doi:10.2471/BLT.19.237487

Figures and Tables

Table 1. Themes and samples of interactions between avatars and patients as outlined by Beaudoin et al. (2021).

Avatar themes	**Samples**	**Patient themes**	**Samples**
Accusations	"You've carried out this task."	Approbation	"Your observation is accurate." "I'm capable of achieving this."
Omnipotence	"I'm feeling scattered everywhere."	Self-deprecation	
Beliefs	"In my opinion, your behavior seems irrational."	Self-appraisal	"I consider myself a kind individual."
Active listening, empathy	"Take your time to unwind, please."	Other beliefs	"You're the one with control over me."
Incitements, orders	"I recommend discontinuing this activity."	Counterattack	"You're responsible for this, not me!"
Coping mechanisms	"Can you explain why my mentioning this makes you sad?"	Maliciousness of the voice	"You seem to be intentionally complicating things for everyone."
Threats	"I'll bring about your downfall."	Negative	"This is quite challenging."
Negative emotions	"Coming to terms with that is challenging for me."	Negation	"I didn't perform this action."
Self-perceptions	"I view myself as being insignificant."	Omnipotence	"I possess unmatched abilities."
Positive emotions	"I'm unparalleled in the entire world."	Disappearance of the voice	"Please disappear!"
Provocation	"Try preventing me from causing you harm."	Positive	"I'm experiencing a positive emotional state."
Reconciliation	"Shall we work towards reconciliation?"	Prevention	"I'll attempt to ignore your presence."

Reinforcement	"Give this another attempt."	Reconciliation of the voice	"Shall we become friends?"
		Self-affirmation	"I am capable of accomplishing this. "

Table 2. Characteristics of sampled participants

Characteristics	**Value (n=18)**
Sex (# male, #female)	16,2
Age (mean in years)	42.6 ± 6.2
Education (mean in years)	12.8 ± 3.6
Ethnicity (Caucasian, others)	94.4%,5.6%
% on Clozapine	61.1%

Table 3. Individual classifiers and ensemble mean scores for accuracy, precision, recall and f1-score for the classification of avatar interactions.

Models	**Accuracy (range)**	**Precision (range)**	**Recall (range)**	**f1-score (range)**
LSVC	0.66 (0.64-0.67)	0.70 (0.69-0.71)	0.66 (0.65-0.67)	0.66 (0.65-0.67)
MultinomialNB	0.48 (0.47-0.48)	0.62 (0.47-0.49)	0.48 (0.47-0.49)	0.42 (0.41-0.43)
MLP	0.66 (0.64-0.67)	0.68 (0.65-0.69)	0.66 (0.65-0.67)	0.66 (0.65-0.67)
XGB	0.54 (0.54-0.55)	0.64 (0.64-0.65)	0.56 (0.56-0.57)	0.56 (0.56-0.57)
KNN	0.57 (0.55-0.58)	0.65 (0.63-0.67)	0.58 (0.56-0.60)	0.56 (0.54-0.58)
Ensemble	0.71 (0.69-0.72)	0.71 (0.69-0.72)	0.71 (0.69-0.72)	0.70 (0.69-0.71)

Table 4. Individual classifiers and ensemble mean scores for accuracy, recall , precision and f1-score for the classification of avatar interactions.

Models	**Accuracy (range)**	**Precision (range)**	**Recall (range)**	**f1-score (range)**
LSVC	0.57 (0.56-0.58)	0.62 (0.60-0.63)	0.57 (0.56-0.58)	0.58 (0.57-0.9)
MultinomialNB	0.44 (0.44-0.45)	0.50 (0.50-0.51)	0.44 (0.43-0.44)	0.40 (0.39-0.41)
MLP	0.54 (0.53-0.55)	0.57 (0.55-0.57)	0.54 (0.53-0.55)	0.55 (0.54-0.56)
XGB	0.48 (0.48-0.49)	0.50 (0.49-0.51)	0.48 (0.48-0.49)	0.49 (0.48-0.50)
KNN	0.45 (0.43-0.46)	0.51 (0.48-0.53)	0.46 (0.45-0.47)	0.46 (0.45-0.47)

Ensemble	0.58 (0.57-0.9)	0.58 (0.57-0.9)	0.58 (0.57-0.9)	0.58 (0.57-0.9)

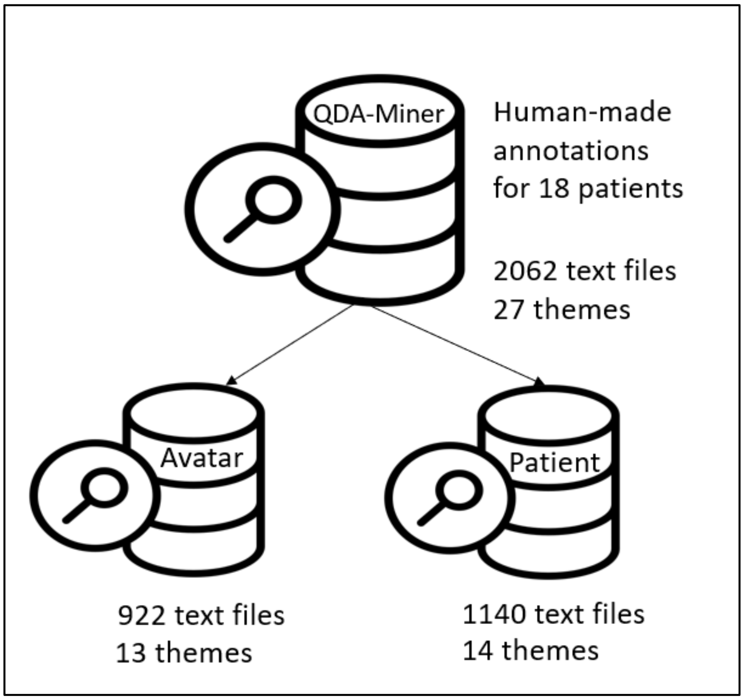

Figure 1. Dataset for the Avatar Therapy corpus

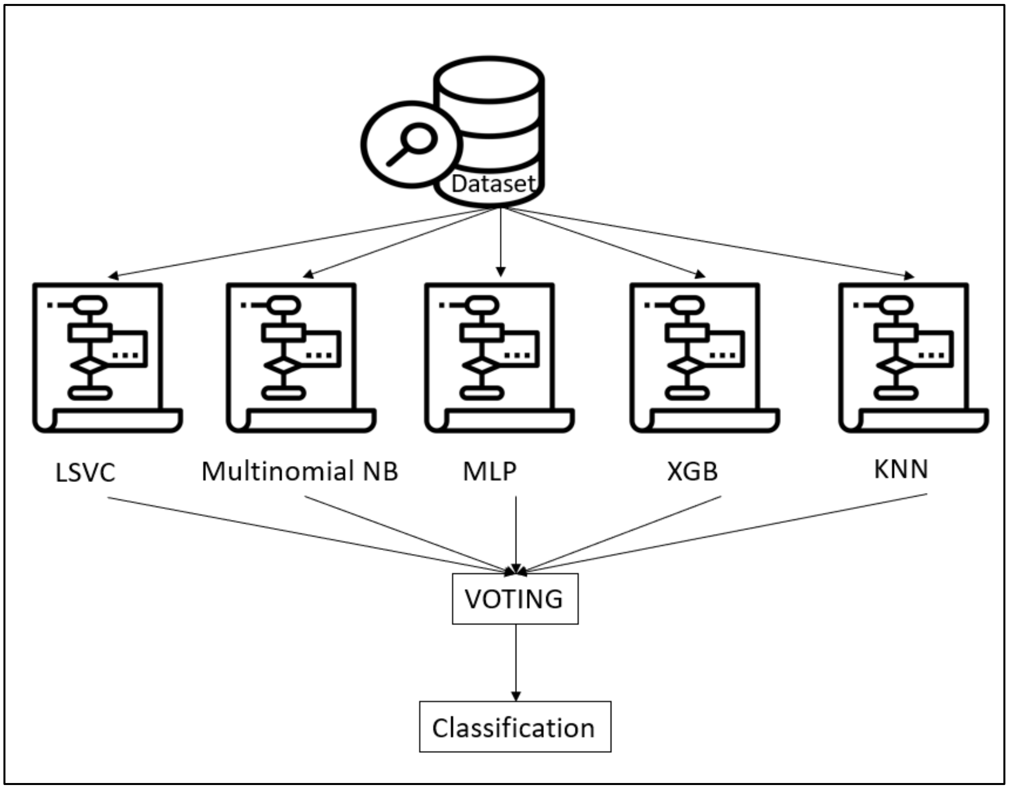

Figure 2. Ensemble modeling is applied to each classification, with this process carried out for both the avatar conceptual dataset and the patient conceptual datasets.

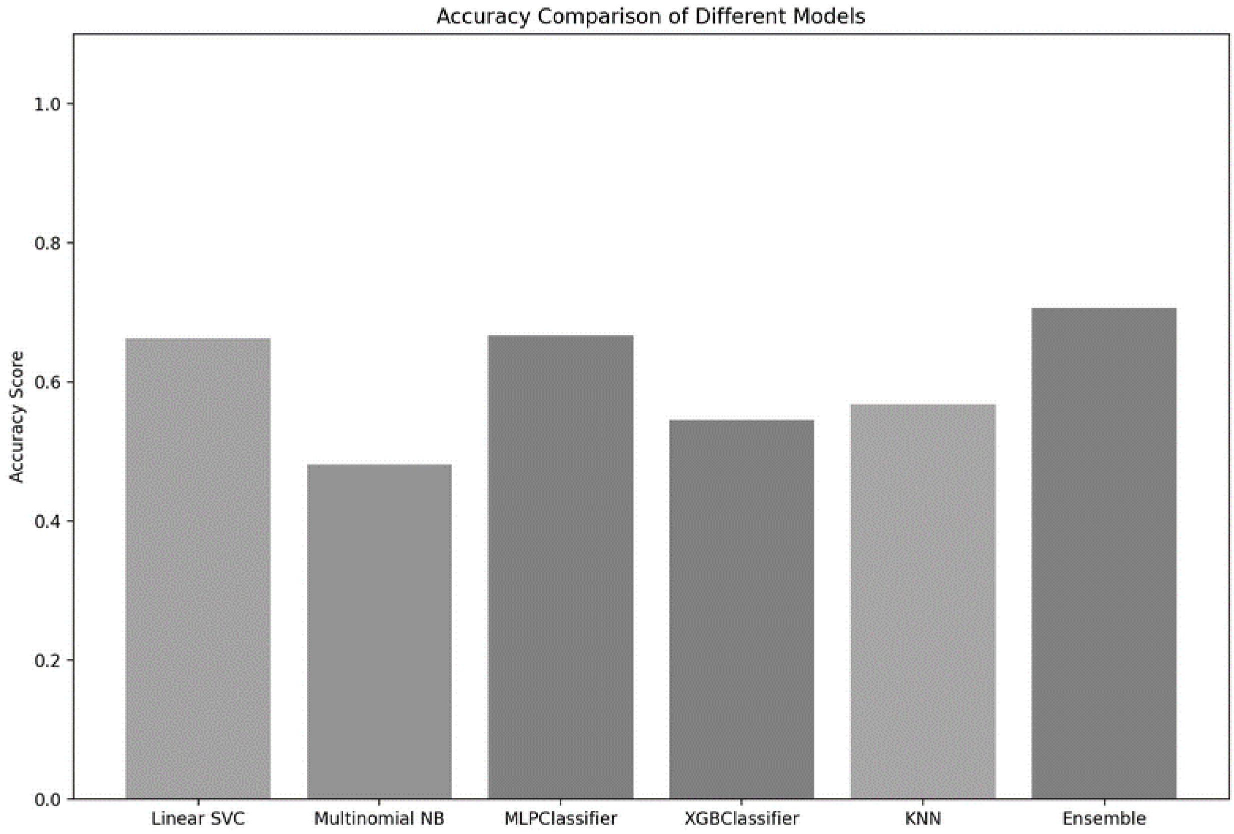

Figure 3. Accuracy comparison of the individual models implemented as well as the ensemble model which encompasses all the individual models' classification of avatar interactions.

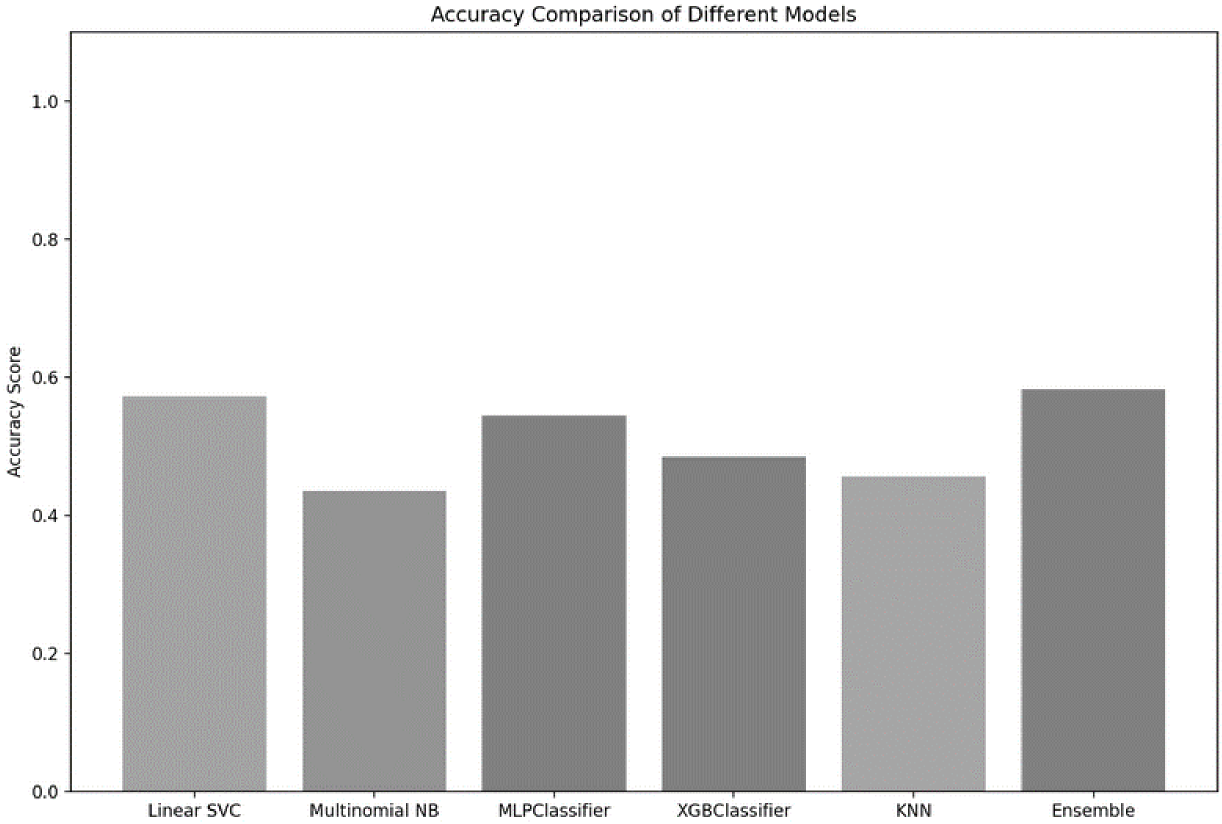

Figure 4. Accuracy comparison of the individual models implemented as well as the ensemble model which encompasses all the individual models for the classification of patient interactions.

Article 8. Enhancing Predictive Power: Integrating a Linear Support Vector Classifier with Logistic Regression for Patient Outcome Prognosis in Virtual Reality Therapy for Treatment-Resistant Schizophrenia

Alexandre Hudon

Mélissa Beaudoin

Kingsada Phraxayavong

Stéphane Potvin

Alexandre Dumais

Publié dans *Journal of Personalized Medicine (2023).*
Article révisé par les pairs.

Abstract

(1) Background: Approximately 30% of schizophrenia patients are known to be treatment-resistant. For these cases, more personalized approaches must be developed. Virtual reality therapeutic approaches such as avatar therapy (AT) are currently undergoing investigations to address these patients’ needs. To further tailor the therapeutic trajectory of patients presenting with this complex presentation of schizophrenia, quantitative insight about the therapeutic process is warranted. The aim of the study is to combine a classification model with a regression model with the aim of predicting the therapeutic outcomes of patients based on the interactions taking place during their first immersive session of virtual reality therapy. (2) Methods: A combination of a Linear Support Vector Classifier and logistic regression was conducted over a dataset comprising 162 verbatims of the immersive sessions of 18 patients who previously underwent AT. As a testing dataset, 17 participants, unknown to the dataset, had their first immersive session presented to the combinatory model to predict their clinical outcome. (3) Results: The model accurately predicted the clinical outcome for 15 out of the 17 participants. Classification of the therapeutic interactions achieved an accuracy of 63%. (4) Conclusion: To our knowledge, this is the first attempt to predict the outcome of psychotherapy patients based on the content of their interactions with their therapist. These results are important as they open the door to personalization of psychotherapy based on quantitative information about the interactions taking place during AT.

Keywords

precision medicine; virtual reality therapy; auditory hallucinations; schizophrenia; machine learning

Introduction

Schizophrenia and Treatment Resistance

A recent study estimated that, in 2019, about 418 million disability-adjusted life years are caused by mental disorders, with a worldwide economic burden evaluated at USD 5 trillion [1]. Amongst mental health disorders, schizophrenia has a relatively small prevalence of less than 1% but significantly contributes to this global burden of mental disorders [2,3]. This psychotic disorder was originally defined by Eugen Bleuler in 1908 [4,5]. Schizophrenia is a functional psychotic condition marked by the presence of delusional beliefs, hallucinations, and disruptions in cognition, perception, and behavior [6,7]. Hallucinations are more often auditory in schizophrenia and are also part of an ensemble of symptoms known as positive symptoms [8,9]. This constellation of symptoms includes hallucinations, delusions, and cognitive impairments. This medical condition is not benign considering that people with schizophrenia often have a life expectancy reduction of 10 to 25 years because they have a greater suicide risk and more physical health issues than the general population [10]. Moreover, there is a higher risk of violence when the symptoms are not addressed, including hetero-aggression, victimization, and self-harm [11,12]. Indeed, violent offenses, including homicides, are also more likely in schizophrenia and other psychotic disorders [13]. However, most of the extra risk appears to be mitigated by concomitant drug usage [14,15,16]. Patients suffering from schizophrenia and benefiting from treatment usually reduce their risk of self-harm and violence [17,18]. They also experience improvements in their quality of life and life expectancy [19].

In schizophrenia, the treatment course usually includes psychopharmacological and psychotherapeutic approaches [20,21]. As dopamine is a relevant neurotransmitter involved in the positive symptoms observed in schizophrenia, the psychopharmacological component most often includes dopaminergic receptor antagonists, also known as antipsychotics [22,23]. However, about 30% of people with schizophrenia will not respond adequately to these medications and will be referred to as treatment-resistant schizophrenia (TRS) patients [24]. For those patients, most clinical guidelines support that a failure to respond to two antipsychotics warrants a trial of clozapine: a second-generation antipsychotic [25]. This approach is the current standard treatment for patients suffering from TRS. However, it is estimated that 40–70% of patients with TRS have persistent symptoms despite clozapine use [26,27]. Such symptoms include persistent auditory hallucinations, which represent the most prevalent and disabling symptoms in schizophrenia. Trials of cognitive behavioral therapy (CBT) have been recommended as an adjunctive approach in the treatment of positive and persistent positive symptoms in TRS [28]. While this form of therapy is an interesting avenue to reduce positive symptoms in people suffering from TRS, results are mitigated, hence why new therapeutic strategies are currently emerging [29].

Current Innovation: Virtual Reality Therapy

Virtual reality therapy (VRT) for patients suffering from TRS is also known as avatar therapy (AT). This psychotherapeutic approach, developed by Julian Leff and his team in 2008, involves the use of an immersive virtual reality system in which TRS patients interact with an avatar, which virtually represents their main persistent auditory verbal hallucination while being controlled and animated by the therapist [30].

Numerous studies have shown that AT is useful in reducing auditory and verbal hallucinations [31,32]. AT is also a psychotherapeutic treatment developed at the Centre de Recherche de Institut universitaire en santé mentale de Montréal (CR-IUSMM). Studies are undergoing to evaluate its efficacy compared to that of CBT [32]. Nine weekly therapy sessions are planned as part of the therapeutic process, most of which are derived from other relational therapeutic approaches and CBT [32]. To accurately depict the patient’s own representation of their most upsetting speech hallucination (persistent auditory hallucinations; “voice”), the therapist and the patient work together to create an avatar during the first session using Unity software. The design of the avatar allows for the consideration of a wide range of traits, including gender, facial features, breadth, voice, and height. Patients then engage with the avatar during a portion of the following eight sessions while wearing a virtual reality headset. The therapist controls the animation of the avatar and operates an external speech modulator system to control their voice modulation.

Pilot studies examining the effects of AT showed that patients with TRS improved by significantly lowering the frequency of persistent auditory hallucinations as well as the distress associated with these symptoms, with large effect sizes [31,32]. Moreover, this therapy also significantly improved their quality of life [33]. To better document and analyze the content of the immersive portions (dialogues between the avatar and the patient), two main qualitative studies were conducted. Five key themes emerged from an initial qualitative investigation of the discourse elements of AT among the patients: emotional responses to voices, beliefs about voices and schizophrenia, self-perceptions, coping mechanisms, and aspirations [34]. Then, a second in-depth investigation by Beaudoin and her team provided further details. The verbatims (immersive session transcripts) of 18 patients who received AT were analyzed for content. Positive techniques (comprising six sub-themes) and confrontational techniques (comprising eight sub-themes) were the two main core interactions identified for the avatar [35]. The patients' self-perceptions, emotional reactions, goals, coping strategies, and beliefs were all highlighted as five major themes comprising 14 sub-themes. Qualitative data are vast and insightful, and they can lead to the generation of novel hypotheses [35]. However, they lacks the quantitative equivalent required to identify the precise components of therapy that may help patients achieve favorable outcomes. Considering the above limitations and potential human biases involved in the classification of verbal interactions, automated classification approaches have been attempted to analyze new verbatims of patients who underwent AT by classifying every interaction into one of the sub-themes identified by Beaudoin and her team. Several machine learning classifiers were compared [36]. In the end, the linear support vector machine classifier (LSVC) performed best for classification of therapeutic interactions taking place in AT [37].

AT being a novel approach for which access is currently limited to research participants, a better understanding of the therapeutic outcomes and predictors of such outcomes could be helpful when assessing which patients will benefit from the therapy. Considering that therapeutic components (i.e., verbal interactions between the patient and their avatar) can be qualitatively and quantitatively assessed, it could be beneficial to assess their predictive power of the reduction of persistent auditory hallucinations and consequently better personalize the treatment of TRS patients.

Precision Medicine Using Predictive Approaches

Modern medicine encompasses several precision medicine avenues, including the use of artificial intelligence to conduct an array of tasks: helping clinicians establish diagnosis, choosing treatment plans, outcome prediction, etc. [38,39,40]. Although precision medicine lags in psychiatry compared to other areas of medicine, it has been highlighted as an avenue to help patients and clinicians in achieving personalization of treatment plans [41]. A recent example is the implementation of a crisis predictor from the exploration of 581 656 medical records for patients suffering from various psychiatric disorders [42]. The model, developed by Garriga and his team, predicted crises with a sensitivity of 58% and a specificity of 85%, achieving an area under the receiver operating characteristic curve of 0.797 and an area under the precision-recall curve of 0.159 [42]. As for treatment responses, a relevant review identified eight studies about patients suffering from depression in which implementations of machine learning models have shown good treatment response prediction (up to 80% accuracy), frequently surpassing usual regression techniques [43]. Although this literature is scarce, several reviews on the application of machine learning to psychiatry and psychology support the idea that this avenue could provide more personalized care for patients [44]. Such use of machine learning in VRT could provide an insight as to which patients are more likely to benefit from this approach earlier or later in their recovery process. Moreover, such tools could also help therapists in planning their immersive sessions more effectively and ultimately conduct better tailored therapeutic sessions to help each patient in achieving favorable outcomes.

Objective and Hypothesis

By combining a classification model with a regression model, the aim of the study is to predict patients’ therapeutic outcome based on the interactions taking place in their first AT therapeutic session. The classification model is used to classify a verbatim into the right interaction themes, and the regression model is used to determine the predictive value of the newly annotated verbatim. Given the prior use of automated classification algorithms in virtual reality therapy, along with the binary nature of the therapeutic outcome, it is hypothesized that this process can be employed to predict a patient's therapeutic outcome. To the best of our knowledge, this is the first attempt to predict the outcome of a patient in psychotherapy based on the content of their interactions with their therapist.

Materials and Methods

Participants and Recruitment

Data from participants involved in previous studies were used for the purpose of this investigation. These individuals were signed up for the clinical trial with the NCT03585127 identifier that was listed on ClinicalTrials.gov (accessed on 12 September 2023) [32]. All of them underwent six to ten one-hour psychotherapy sessions, eight of which involved interaction with an avatar that represented their auditory verbal hallucinations, while the creation of the avatar took place during the first session. The immersive portion of the therapeutic sessions between the patient and the avatar lasted between 15 and 50 min. Recruitment occurred at the CR-IUSMM between 2017 and 2022; participants were either referred by their treating team or self-referred. Inclusion criteria included an age of 18 or older and a diagnosis of TRS, characterized by a persistent auditory hallucination despite two or more trials of dopaminergic antagonists. The dataset consisted of treatment interventions for 18 patients, and the prediction power was tested using data from 17 other participants not included in the initial dataset. Therefore, for these 17 patients, the therapeutic outcome was known to the therapists but was unknown to the predictive algorithm.

Dataset: Corpus of Avatar Therapy and Features

A dataset comprising a total of 162 handwritten treatment transcripts of 18 patients who received VRT between 2017 and 2020 at our institution, corresponding to up to 10 therapy sessions per patient, was developed. The transcripts were written in Canadian French. The 27 themes listed in Beaudoin et al. 2021 were used to hand annotate transcripts [35]. This qualitative analysis of AT was carried out in previous research. Every one of the distinct interactions was individually coded by two study assistants. The same two research assistants cross-validated the robustness of the coding grid. Annotations were performed using the qualitative data analysis program QDA Miner version 5 (Provalis Research) [45]. Then, these were retrieved as text files from QDA Miner and categorized under two conceptual databases, Avatar and Patient, in order to optimize the automatic categorization. These text files contained between one and forty interactions of the same topic. The conceptual datasets were created in accordance with Figure 1's representation.

-- Please insert Figure 1 here --

-- Please insert Table 1 here --

Overview of the Predictive Approach

This study combines an automated classification algorithm with a logistic regression. The classification model was trained based on the AT dataset for each conceptual database. Then, an unannotated verbatim for the first immersive session of a participant who previously underwent AT (but unknown to the AT dataset) was presented to the classification model. Once automated classification of each interaction between the avatar and the patient was achieved, the frequency for each interaction was compiled and passed through the logistic regression model. A prediction of the outcome could then be achieved. The overall flow of this predictive approach is presented in Figure 2.

-- Please insert Figure 2 here --

Automated Classification of Verbatims

Previous Work

A support vector machine (SVM) was implemented as per previous work on avatar therapy [37]. This machine learning method is used for both regression and classification problems. It operates by identifying the ideal hyperplane in a high-dimensional feature space that best divides several classes [46]. This hyperplane is set up to maximize the distance (margin) between the classes, which enhances the model's ability to generalize to new data. SVMs are a well-liked option for text classification jobs because of their capacity to handle high-dimensional, sparse, and non-linear data [46].

In this study, a linear form of SVM was combined to a term frequency-inverse document frequency statistic (TF-IDF). Compared to various SVMs with a tokenizer combination, a TF-IDF performs best with text categorization [47]. The TfidfVectorizer class, available in the Scitkit-Learn open library, was chosen for the TF-IDF tokenization as it allows for the conversion of raw text (extracted interactions from interview transcripts) into numerical vectors [48]. Stop-words can be accounted for by customizing vectorizers. The features were expected to be linearly separable since the classification categories were created so that text entities would be divided based on their inherent qualities, which are fundamentally distinct and specified in Beaudoin et al. (2021). A GridSearchCV (GSCV) was employed to guarantee the LSVC algorithm's optimal performance and to improve search tactics [48]. The benefit of a GSCV is that it allows the user to test for various hyper-parameters and cross-validate the LSVC's classification to find the optimal set of LSVC parameters and TfidVectorizer parameter variables. The full implementation details, including implementation parameters and hypertuning, can be found in Hudon et al. [37].

Linear Support Vector Classifier

The LSVC employs a linear kernel as opposed to regular SVM [49]. A kernel is a mathematical function that transforms data into a higher-dimensional feature space and is used in several machine learning techniques [50]. Kernels are crucial to algorithms' capacity to solve complex problems that might be difficult or even impossible to handle in the original input space [50]. When the data can be separated linearly, a linear kernel is thus applied. Scikit-Learn's SVC class of the SVM library, which has the specification to use a linear kernel, is the implementation of the SVC used in this work [48].

Prediction of Patient's Outcome

In this study, the outcome was measured based on the change in auditory hallucinations, measured using the Psychotic Symptom Rating Scales (PSYRATS) auditory hallucinations subscale [51]. The PSYRATS is a clinical evaluation instrument used to gauge how severe and specific psychotic symptoms are in psychotic individuals. The auditory hallucinations subscale is comprised of 11 items: frequency, duration, controllability, loudness, location, amount and degree of negative content, severity and intensity of distress, beliefs about the origin of voices, and disruption [51]. A participant who experienced a decrease of 20% in the PSYRATS auditory hallucination subscale was defined as being a good responder, whereas other participants were referred to as non-responders.

Data Analysis and Validation

Training and Cross-Validation

For each conceptual database, a partitioning method was implemented, with 70% of the annotated documents being used to train the LSVC and the remaining 30% being used for testing. The goal was to determine a statistical likelihood for the LSVC, represented by a classification predictive score, which would indicate how well an interaction might be classified. To follow suggested design practices, the training and testing sets were purposefully kept apart [52]. The predictive score reflects the average accuracy as determined by the F1-Score. The K-Fold model from the Scikit-Learn module was used to build a tenfold cross-validation strategy for both the logistic regression algorithm and the linear support vector algorithm.

Classification Analysis

Information on the classification performance of each topic, including accuracy, recall, and F1-Score for each method, was gathered using the Classification Report tool from the Scikit-Learn metrics module [48]. The F1-Score depicts the accuracy of theme categorization, recall of the sensitivity of the prediction, and precision of the positive predictive value. To provide a comprehensive evaluation of classification accuracy, the F1-Score, a widely used metric in text classification, finds a compromise between precision and recall [53]. Therefore, the harmonic mean of recall and accuracy is the F1-Score [53].

Prediction Analysis

Considering the binary outcome of virtual reality therapy (good responder versus non-responders), a logistic regression was implemented. The interactions between avatar and the patient as defined above were used as features to determine the outcome of the regression. The LogisticRegressionCV class, which is a logistic regression with build-in cross-validation from the Scikit-Learn library, was used [48]. The adjusted R2 score was representative of our predictive score. A score of 1 would indicate that the model explains all the variation of the dependent variable around its mean compared to a score of 0, which means that the model does not explain at all the observed variations. Collinearity between the different variables was accounted for in the logistic regression algorithm by providing the variance inflation factor (VIF) for each feature. To build the logistic regression model, the average frequency for each type of interaction for the first set of participants was used to construct a second dataset used for predictive purposes. This dataset contained all the frequencies of the interaction themes of the 18 participants who previously completed AT. Considering that the interaction themes are the features of the predictive model, this approach is needed to predict the potential therapeutic outcome of a newly annotated verbatim. The data used for this are available in Supplementary Material S1.

Finally, to predict patients' outcomes, automatically annotated verbatims of the first immersive session of 17 participants (second set of participants) were used. The frequency of each type of interaction was calculated and used by the model to conclude if the participant was forecasted to be a good responder or a non-responder. Statistical significance is defined by a likelihood ratio p-value smaller than 0.05 [54].

Results

Sample Characteristics

Interactions taking place in the verbatims of 18 patients were used to construct the initial interaction dataset, from which interaction frequencies were used to make the prediction of the outcome. The characteristics of the sampled patients can be found in Table 2.

--Please insert Table 2 here--

Performance of the Classification Algorithm

The average performance of the LSVC for the automatic annotation of the verbatims of the second set of participants can be found in Table 3. The precision score ranges from 0.62 to 0.67, the recall ranges from 0.58 to 0.65, and the F1-Score ranges from 0.60 to 0.65. Classification scores for participants 007 and 016 are the lowest, whereas the average accuracy score for annotation is 63% (as per the F1-Score). The classification of avatar themes performed better than patient interaction themes (70% versus 62%). Sample performances for each theme and class balances for each feature are found in Table S2.

-- Please insert Table 3 here--

-- Please insert Table S2 here--

The logistic regression model achieved an adjusted R2 performance of 0.736 with a likelihood p-value of 0.04. From the first immersive session verbatim of the 17 participants, a total of 15 had a predicted outcome corresponding to their true outcome (88.2% accurate predictions). Errors occurred for participants 003 and 016. Amongst the previously identified 27 themes, 16 were selected for the model when accounting for collinearity and relevancy. Coefficients of the features as well as model performances are found in Table A1. Logistic regression score and outcomes are presented for each participants in Table 4.

-- Please insert Table A1 here--

-- Please insert Table 4 here --

Discussion

This study aimed to combine a classification model with a regression model to predict patients' treatment outcome based on interactions in their first AT immersive session. The verbatims (first immersive session) of 17 participants who previously underwent AT were automatically annotated based on a corpus from previous participants of AT, and a prediction was effectuated based on the frequency of each type of interaction that took place during the first VR session. As a result, the combination of the models predicted accurately the outcome of 15 out of 17 participants with an average accuracy score for the automated annotations of 63%.

Prediction of psychotherapeutic outcome is a potentially interesting avenue to personalize treatments for patients suffering from severe mental illnesses. However, prior works on the topic of predicting patient outcomes using patient risk factors and demographics were mostly inconclusive [55,56]. Patient factors alone cannot be the sole elements to predict patients' outcomes as therapeutic processes are highly dependent on the therapeutic alliances and patient–therapist interactions. A five-year longitudinal study protocol highlighted the importance of considering the therapist's interpersonal skills in relation with the outcome of the psychotherapy [57]. Other elements of the psychotherapeutic process are also in the process of being modeled. For example, a recent study trained several machine learning algorithms on patients' self-reported side effects of psychotherapy to predict performances of the psychotherapeutic outcomes. They achieved an accuracy of 79.7% using Random Forest-based machine learning classifiers [58]. However, such implementation of prediction classifiers is to be used with caution as numerous biases might be implied such as class imbalances and the definition of the therapeutic outcome in their context of relevance. When compared to other fields of medicine, where terms (e.g., signs and symptoms) may be used to help categorization, psychotherapy such as AT often utilizes a larger range of words and contextual sentences. This might explain why the automated classification does not reach perfect accuracy.

As for the predictive performance, few studies address predictive indicators of therapeutic outcomes based on therapeutic interactions. However, a recent literature review on the potential uses of machine learning to predict responses to CBT for different mental health disorders identified algorithm performances ranging from 67.3% up to 87.0% [59]. The performance of our model (88.2%) may differ from this range of performances because AT differs from CBT and uses a protocolized approach for patients suffering from TRS. However, the complexity of TRS and the variety of interactions between the therapists and the patients might account for the lower classification scores which is often observed when the elements of the corpus are overlapping or imbalanced. Another explanation could be that AT is based on an important role-playing relationship between the therapist and the patient, necessarily less linear in its approach, which could explain the low classification accuracy.

Limitations

The performance trend for the LSVC is to be re-evaluated when additional patients are added to the dataset. Of note, transcripts used for this study's analysis were written in Canadian French, and consequently, finding vectorizers that included stop-words particularly for the Canadian French language was challenging. Stop-words are terms that are often not included in the tokenization process because their meanings are either vague or unimportant. The accuracy of the analysis could have been impacted by the lack of suitable stop-words for Canadian French, which could lead to irrelevant terms being included in the word vectors and distorting the final findings. Finally, it is of importance to note that the performance of the regression algorithm was based on a limited dataset which might affect its performance and is dependent of the classification algorithm.

Conclusions

To conclude, this study demonstrated that classification algorithms such as LSVC can be combined with a predictive algorithm to predict patients' outcomes for AT. Out of 17 participants unknown to the original dataset, the outcomes of 15 were accurately predicted based on the frequency of their interactions during their first immersive session. Automated classifications of the interactions taking place during the immersive session achieved performances comparable to previous studies on the subject. These results present an interesting avenue for the personalization of patients’ experiences with AT as the therapist might use this insight between the immersive sessions to help prepare their next sessions to enhance patients’ likelihood of achieving a favorable outcome. This could be carried out by identifying interactions linked to a positive outcome and encouraging such interactions. To the best of our knowledge, this is the first implementation of a predictor based on content elements of the therapeutic process. This opens the door to future studies to explore the possibility of using such classifiers in different psychotherapeutic contexts or by mixing potential predictive elements such as emotional content, therapeutic alliance, and patients’ characteristics.

Supplementary Materials

The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jpm13121660/s1, Table S1: Frequencies of all avatar and patient themes across avatar therapy; Table S2: Sample performances for each theme and class balances for each feature.

Author Contributions

Conceptualization, A.H., M.B., K.P., S.P. and A.D.; methodology, A.H. and A.D.; validation, A.H. and A.D.; formal analysis, A.H.; investigation, A.H.; data curation, A.H.; writing—original draft preparation, A.H. and M.B.; writing—review and editing, A.H., M.B., K.P., S.P. and A.D.; supervision, K.P., S.P. and A.D.; project administration, K.P.; funding acquisition, K.P., S.P. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work was indirectly supported indirectly by Le Fonds de recherche du Québec—Santé (FRQS); Otsuka Canada Pharmaceutical Inc.; Chaire Eli Lilly Canada de recherche en schizophrénie; MEI (Ministère de l’Économie et de l’Innovation); Services et recherches psychiatriques AD; Fonds d’excellence en recherche Apogée Canada. These fundings bodies had no part in the data

collection, analysis, interpretation of data, and in writing the manuscript.

Institutional Review Board Statement

This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients. Patients that are part of this study were selected based on the proof-of-concept trial from Percy du Sert's 2018 study and Dellazizzo's 2021 study. The trial was conducted in accordance with the Declaration of Helsinki and was approved by the institutional ethical committee (CER IPPM 16-17-06). We obtained written informed consent from all patients.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The datasets generated and/or analyzed during the current study containing patients' verbatims are not publicly available due to patients' privacy but are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedi-cine. 2022;54:101675. Published 2022 Sep 28. doi:10.1016/j.eclinm.2022.101675

2. Janoutová J, Janácková P, Serý O, et al. Epidemiology and risk factors of schizophrenia. Neuro Endocrinol Lett. 2016;37(1):1-8.

3. Mueser KT, McGurk SR. Schizophrenia. Lancet. 2004;363(9426):2063-2072. doi:10.1016/S0140-6736(04)16458-1

4. Ashok AH, Baugh J, Yeragani VK. Paul Eugen Bleuler and the origin of the term schizophrenia (SCHIZOPRENIEGRUPPE). Indian J Psychiatry. 2012;54(1):95-96. doi:10.4103/0019-5545.94660

5. Orsolini L, Pompili S, Volpe U. Schizophrenia: A Narrative Review of Etiopathogenetic, Diagnostic and Treatment Aspects. Journal of Clinical Medicine. 2022; 11(17):5040. https://doi.org/10.3390/jcm11175040

6. Stępnicki P, Kondej M, Kaczor AA. Current Concepts and Treatments of Schizophrenia. Molecules. 2018; 23(8):2087. https://doi.org/10.3390/molecules23082087

7. Humpston CS, Broome MR. Thinking, believing, and hallucinating self in schizophrenia. Lancet Psychiatry. 2020;7(7):638-646. doi:10.1016/S2215-0366(20)30007-9

8. Lim A, Hoek HW, Deen ML, Blom JD; GROUP Investigators. Prevalence and classification of hallucinations in multiple sensory modalities in schizophrenia spectrum disorders. Schizophr Res. 2016;176(2-3):493-499. doi:10.1016/j.schres.2016.06.010

9. Montagnese M, Leptourgos P, Fernyhough C, et al. A Review of Multimodal Hallucinations: Categorization, Assessment, Theoretical Perspectives, and Clinical Recommendations. Schizophr Bull. 2021;47(1):237-248. doi:10.1093/schbul/sbaa101

10. Laursen TM, Munk-Olsen T, Vestergaard M. Life expectancy and cardiovascular mortality in persons with schizophrenia. Curr Opin Psychiatry. 2012;25(2):83-88. doi:10.1097/YCO.0b013e32835035ca

11. Girasek H, Nagy VA, Fekete S, Ungvari GS, Gazdag G. Prevalence and correlates of aggressive behavior in psychiatric inpatient populations. World J Psychiatry. 2022;12(1):1-23. Published 2022 Jan 19. doi:10.5498/wjp.v12.i1.1

12. Cho W, Shin WS, An I, Bang M, Cho DY, Lee SH. Biological Aspects of Aggression and Violence in Schizophrenia. Clin Psychopharmacol Neurosci. 2019;17(4):475-486. doi:10.9758/cpn.2019.17.4.475

13. Tiihonen J, Isohanni M, Räsänen P, Koiranen M, Moring J. Specific major mental disorders and criminality: a 26-year pro-spective study of the 1966 northern Finland birth cohort. Am J

Psychiatry. 1997;154(6):840-845. doi:10.1176/ajp.154.6.840

14. Manseau M, Bogenschutz M. Substance Use Disorders and Schizophrenia. Focus (Am Psychiatr Publ). 2016;14(3):333-342. doi:10.1176/appi.focus.20160008

15. Hudon A, Dellazizzo L, Phraxayavong K, Potvin S, Dumais A. Association Between Cannabis and Violence in Communi-ty-Dwelling Patients With Severe Mental Disorders: A Cross-sectional Study Using Machine Learning. J Nerv Ment Dis. 2023;211(2):88-94. doi:10.1097/NMD.0000000000001604

16. Fazel S, Långström N, Hjern A, Grann M, Lichtenstein P. Schizophrenia, substance abuse, and violent crime. JAMA. 2009;301(19):2016-2023. doi:10.1001/jama.2009.675

17. Wimberley T, MacCabe JH, Laursen TM, et al. Mortality and Self-Harm in Association With Clozapine in Treatment-Resistant Schizophrenia. Am J Psychiatry. 2017;174(10):990-998. doi:10.1176/appi.ajp.2017.16091097

18. Kasckow J, Felmet K, Zisook S. Managing suicide risk in patients with schizophrenia. CNS Drugs. 2011;25(2):129-143. doi:10.2165/11586450-000000000-00000

19. Guo X, Zhang Z, Zhai J, et al. Effects of antipsychotic medications on quality of life and psychosocial functioning in patients with early-stage schizophrenia: 1-year follow-up naturalistic study. Compr Psychiatry. 2012;53(7):1006-1012. doi:10.1016/j.comppsych.2012.03.003

20. Kokurcan A, Güriz SO, Karadağ H, Erdi F, Örsel S. Treatment strategies in management of schizophrenia patients with persistent symptoms in daily practice: a retrospective study. Int J Psychiatry Clin Pract. 2021;25(3):238-244. doi:10.1080/13651501.2021.1879157

21. National Collaborating Centre for Mental Health (UK). Psychosis and Schizophrenia in Adults: Treatment and Management. London: National Institute for Health and Care Excellence (UK); 2014.

22. Kesby JP, Eyles DW, McGrath JJ, Scott JG. Dopamine, psychosis and schizophrenia: the widening gap between basic and clinical neuroscience. Transl Psychiatry. 2018;8(1):30. Published 2018 Jan 31. doi:10.1038/s41398-017-0071-9

23. Novak, G.; Seeman, M.V. Dopamine, Psychosis, and Symptom Fluctuation: A Narrative Review. Healthcare 2022, 10, 1713. https://doi.org/10.3390/healthcare10091713

24. Patel KR, Cherian J, Gohil K, Atkinson D. Schizophrenia: overview and treatment options.

P T. 2014;39(9):638-645.

25. Bittner RA, Reif A, Qubad M. The ever-growing case for clozapine in the treatment of schizophrenia: an obligation for psychiatrists and psychiatry. Curr Opin Psychiatry. 2023;36(4):327-336. doi:10.1097/YCO.0000000000000871

26. Chakrabarti S. Clozapine resistant schizophrenia: Newer avenues of management. World J Psychiatry. 2021;11(8):429-448. Published 2021 Aug 19. doi:10.5498/wjp.v11.i8.429

27. Shah P, Iwata Y, Brown EE, et al. Clozapine response trajectories and predictors of non-response in treatment-resistant schizophrenia: a chart review study. Eur Arch Psychiatry Clin Neurosci. 2020;270(1):11-22. doi:10.1007/s00406-019-01053-6

28. Ryan M, Sattenspiel D, Chianese A, Rice H. CE: Original Research: Cognitive Behavioral Therapy for Symptom Management in Treatment-Resistant Schizophrenia. Am J Nurs. 2022;122(8):24-33. doi:10.1097/01.NAJ.0000854488.48801.59

29. Morrison AP, Pyle M, Gumley A, et al. Cognitive behavioural therapy in clozapine-resistant schizophrenia (FOCUS): an assessor-blinded, randomised controlled trial. Lancet Psychiatry. 2018;5(8):633-643. doi:10.1016/S2215-0366(18)30184-6

30. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff AP. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work?. Psychosis. 2014;6(2):166-176. doi:10.1080/17522439.2013.773457

31. Craig TK, Rus-Calafell M, Ward T, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial [published correction appears in Lancet Psychiatry. 2017 Nov 29;:]. Lancet Psychiatry. 2018;5(1):31-40. doi:10.1016/S2215-0366(17)30427-3

32. Dellazizzo L, Potvin S, Phraxayavong K, Dumais A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. NPJ Schizophr. 2021;7(1):9. Published 2021 Feb 12. doi:10.1038/s41537-021-00139-2

33. Beaudoin M, Potvin S, Phraxayavong K, Dumais A. Changes in Quality of Life in Treatment-Resistant Schizophrenia Patients Undergoing Avatar Therapy: A Content Analysis. J Pers Med. 2023;13(3):522. Published 2023 Mar 14. doi:10.3390/jpm13030522

34. Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychother. 2018;25(6):878-885. doi:10.1002/cpp.2322

35. Beaudoin M, Potvin S, Machalani A, et al. The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clin Psychol Psychother. 2021;28(3):500-518. doi:10.1002/cpp.2556

36. Hudon A, Phraxayavong K, Potvin S, Dumais A. Comparing the Performance of Machine Learning Algorithms in the Au-tomatic Classification of Psychotherapeutic Interactions in Avatar Therapy. Machine Learning and Knowledge Extraction. 2023; 5(3):1119-1131. https://doi.org/10.3390/make5030057

37. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach. Health Informatics J. 2022;28(4):14604582221142442. doi:10.1177/14604582221142442

38. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-38. doi:10.1038/s41591-021-01614-0

39. Al Kuwaiti A, Nazer K, Al-Reedy A, Al-Shehri S, Al-Muhanna A, Subbarayalu AV, Al Muhanna D, Al-Muhanna FA. A Review of the Role of Artificial Intelligence in Healthcare. Journal of Personalized Medicine. 2023; 13(6):951. https://doi.org/10.3390/jpm13060951

40. Kitsios F, Kamariotou M, Syngelakis AI, Talias MA. Recent Advances of Artificial Intelligence in Healthcare: A Systematic Literature Review. Applied Sciences. 2023; 13(13):7479. https://doi.org/10.3390/app13137479

41. Fakhoury M. Artificial Intelligence in Psychiatry. Adv Exp Med Biol. 2019;1192:119-125. doi:10.1007/978-981-32-9721-0_6

42. Garriga R, Mas J, Abraha S, et al. Machine learning model to predict mental health crises from electronic health records. Nat Med. 2022;28(6):1240-1248. doi:10.1038/s41591-022-01811-5

43. Sajjadian M, Lam RW, Milev R, et al. Machine learning in the prediction of depression

treatment outcomes: a systematic review and meta-analysis. Psychol Med. 2021;51(16):2742-2751. doi:10.1017/S0033291721003871

44. Chen ZS, Kulkarni PP, Galatzer-Levy IR, Bigio B, Nasca C, Zhang Y. Modern views of machine learning for precision psy-chiatry. Patterns (N Y). 2022;3(11):100602. Published 2022 Nov 11. doi:10.1016/j.patter.2022.100602

45. QDA Miner. (Version 5). (2016). Provalis Research.

46. Ben-Hur A, Weston J. A user's guide to support vector machines. Methods Mol Biol. 2010;609:223-239. doi:10.1007/978-1-60327-241-4_13

47. Busagala LSP, Ohyama W, Wakabayashi T, Kimura F: Multiple feature-classifier combination in automated text classification. In Proceeding of 2012 10th IAPR International Workshop on Document Analysis Systems (DAS). Gold Cost, QLD, Australia; March 2012:43-47.

48. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011;12:2825-30.

49. Longato E, Acciaroli G, Facchinetti A, Maran A, Sparacino G. Simple Linear Support Vector Machine Classifier Can Dis-tinguish Impaired Glucose Tolerance Versus Type 2 Diabetes Using a Reduced Set of CGM-Based Glycemic Variability Indices. J Diabetes Sci Technol. 2020;14(2):297-302. doi:10.1177/1932296819838856

50. Müller KR, Mika S, Rätsch G, Tsuda K, Schölkopf B. An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw. 2001;12(2):181-201. doi:10.1109/72.914517

51. Woodward TS, Jung K, Hwang H, et al. Symptom dimensions of the psychotic symptom rating scales in psychosis: a multisite study. Schizophr Bull. 2014;40 Suppl 4(Suppl 4):S265-S274. doi:10.1093/schbul/sbu014

52. Wei Q, Dunbrack RL Jr. The role of balanced training and testing data sets for binary classifiers in bioinformatics. PLoS One. 2013;8(7):e67863. Published 2013 Jul 9. doi:10.1371/journal.pone.0067863

53. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022;12(1):5979. Published 2022 Apr 8. doi:10.1038/s41598-022-09954-8

54. Riedle B, Neath AA, Cavanaugh JE. Reconceptualizing the p-value from a likelihood ratio

test: a probabilistic pairwise comparison of models based on Kullback-Leibler discrepancy measures. J Appl Stat. 2020;47(13-15):2582-2609. Published 2020 Apr 23. doi:10.1080/02664763.2020.1754360

55. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94-98. doi:10.7861/futurehosp.6-2-94

56. Shamout F, Zhu T, Clifton DA. Machine Learning for Clinical Outcome Prediction. IEEE Rev Biomed Eng. 2021;14:116-126. doi:10.1109/RBME.2020.3007816

57. Schöttke H, Flückiger C, Goldberg SB, Eversmann J, Lange J. Predicting psychotherapy outcome based on therapist inter-personal skills: A five-year longitudinal study of a therapist assessment protocol. Psychother Res. 2017;27(6):642-652. doi:10.1080/10503307.2015.1125546

58. Yao L, Xu Z, Zhao X, et al. Therapists and psychotherapy side effects in China: A machine learning-based study. Heliyon. 2022;8(11):e11821. Published 2022 Nov 24. doi:10.1016/j.heliyon.2022.e11821

59. Vieira S, Liang X, Guiomar R, Mechelli A. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. Clin Psychol Rev. 2022;97:102193. doi:10.1016/j.cpr.2022.102193

Figures and Tables

Table 1. Summary of interactions for the Avatar and the Patients in VRT as defined by Beaudoin et al.

Category	**Theme**	**Definition**	**Verbatim Example (translated from French to English)**
Avatar themes			
Confrontational techniques	Threats	The Avatar expresses an intention that is detrimental to the patient.	“I will haunt your family”
	Accusations	The Avatar accuses the patient of having done or thought something. Can also include comments with an accusatory connotation.	“You are always complicating everything”
	Affirmations of omnipotence	The avatar affirms its existence or its omnipotence.	“I am the strongest”
	Incitements, orders	The Avatar encourages or orders the patient to take actions or to think in a certain way.	“Go play with the machines, go spend all your money”
	Provocation	Any words of the avatar which can lead the patient to react, including but not limited to insults, belittlement, and irony. Generally corresponds to what the patient reports to hear on a daily basis.	“I see that you have not gained any self-confidence”
	Manipulation using positive emotions	The Avatar expresses a positive emotion or suggests one in a way that leaves no room for interpretation.	“I am having fun with you” (in response to: “I do not want anything to do with you because you are mean”)
	Manipulation using negative emotions	The Avatar expresses a negative emotion or suggests one in a way that leaves no room for interpretation.	“I am not happy you are telling me to go way”
Positive techniques	Reinforcement	The Avatar encourages the patient to assert themselves or use coping mechanisms,	“You did that all alone, too” (in response to: “I listened to the tapes again, and I think I was courageous”)

		or acknowledges that the patient is starting to do so.	
	Reconciliation	The Avatar suggests that the patient should make peace with them, or agrees to do so.	“Do we make peace?”
	Questions about self-perceptions	The Avatar questions the patient about their perceptions of themselves (e.g., qualities)	“How do you explain that your father finds you courageous?”
	Questions about coping mechanisms	The avatar questions the patient about their coping mechanisms (self-defense, prevention, wishes), or makes comments leading the patient to question their coping mechanisms.	“How could you make me leave? How could you make sure I will not come back?”
	Questions about beliefs	The avatar questions the patient about their beliefs (origin of the voice, intentions of the voice, mental illness), or makes comments leading the patient to question their beliefs.	“I am simply repeating what you think of yourself”
	Empathetic listening, empathy	The Avatar makes empathetic remarks, ask questions, agrees with what the patient says, or reformulates the patient’s words.	“Do you have an example?”
Patient themes			
Emotional response	Positive reactions	The patient expresses a positive emotion while talking with the Avatar.	“I am happy”
	Negative reactions	The patient expresses a negative emotion while talking with the Avatar.	“I cry almost every day because I am sick of this”
Beliefs	Maliciousness of the voice	The patient attributes malicious intents to the Avatar.	“You are telling false stories”
	Omnipotence	The patient states that the voice is omnipotent,	“I have always been sick because of you”

		powerful, or beyond their control.	
	Other beliefs	The patient expresses beliefs about the origin of the voice or about their mental illness.	"I do not think you can do that because you are not real"
Self-perceptions	Self-appraisal	The patient makes self-enhancing comments about themselves.	"You just told me I am a liar, but that is not true, I am honest"
	Self-deprecation	The patient makes demeaning comments about themselves.	"I have qualities, but there are not many qualities, they are rare…"
Coping mechanisms	Self-affirmation	The patient is assertive and demonstrates good self-confidence, usually in response to an affirmation from the Avatar.	"I am not giving you a choice"
	Negation	The patient disagrees with an affirmation from the Avatar	"You will not have control over me"
	Counter-attack	The patient is countering an attack from the Avatar.	"It's getting hot, but it's going to get even hotter if you stay in my head, you're going to get hotter. It's not me who's going to make you burn, it's the inner strength of my good will that's going to make you burn."
	Approbation	The patient agrees with an attack from the Avatar.	"Well, well, alright" (in response to: things are going to go badly for you")
Aspirations	Reconciliation with the voice	The patient expresses their wish to make peace with the avatar, for example by offering to make a deal.	"In the evening, you can come and talk to me. But during the day, leave me alone."
	Disappearance of the voice	The patient expresses their wish that the voice or Avatar disappears, go away, or leave them alone.	"Get out of my head"
	Prevention strategies	The patient establishes cognitive, behavioral, or affective strategies to cope with the distress associated with the voice.	"I will feel OK and I will think of the good god of love that I love on earth"

Table 2. Characteristics of sampled patients for the first set participant included in previously published studies. N=18.

Characteristics	**Value (N=18)**
Male sex (N)	15 (83%)
Age (mean ± SD in years)	40.8 ± 12.1
Education (mean ± SD in years)	13.2 ± 3.4
Caucasian ethnicity (N)	16 (89%)
Clozapine use (N)	11 (61%)

Table 3. Average performances of each participant on the Avatar conceptual database for the metrics. N=17.

Participant #	**Precision**	**Recall**	**F1-Score**
Participant 001	0.65	0.63	0.63
Participant 002	0.67	0.65	0.65
Participant 003	0.63	0.61	0.61
Participant 004	0.65	0.63	0.63
Participant 005	0.65	0.63	0.63
Participant 006	0.64	0.63	0.63
Participant 007	0.62	0.58	0.60
Participant 008	0.65	0.65	0.65
Participant 009	0.65	0.63	0.63
Participant 010	0.65	0.63	0.63
Participant 011	0.64	0.63	0.63
Participant 012	0.64	0.61	0.62
Participant 013	0.65	0.63	0.63
Participant 014	0.65	0.63	0.63
Participant 015	0.65	0.63	0.63
Participant 016	0.62	0.58	0.60
Participant 017	0.65	0.63	0.63
Average scores	0.65	0.63	0.63

Table 4. Comparisons of the true outcome to the predicted outcome for the verbatim of the first immersive session of the 17 participants.

Participant #	**True outcome**	**Predicted outcome**	**True logistic regression score**
Participant 001	Good responder	Good responder	1
Participant 002	Good responder	Good responder	1
Participant 003	Good responder	Good responder	1
Participant 004	Non-responder	Good responder	0
Participant 005	Non-responder	Non-responder	0
Participant 006	Non-responder	Non-responder	0
Participant 007	Non-responder	Non-responder	0
Participant 008	Good responder	Good responder	1
Participant 009	Non-responder	Non-responder	0
Participant 010	Good responder	Good responder	1
Participant 011	Good responder	Good responder	1
Participant 012	Non-responder	Non-responder	0
Participant 013	Good responder	Good responder	1
Participant 014	Good responder	Good responder	1
Participant 015	Non-responder	Non-responder	0
Participant 016	Non-responder	Good responder	0
Participant 017	Non-responder	Non-responder	0

Table A1. Logistic regression model's performances

Model	**Logistic regression**
Pseudo R-Squared	0.736
Likelihood ration p-value	0.04
intercept	-4.651781763166642
coefficients:	
Self-appraisal (Patient)	0.04883651593641261
Self-affirmation (Patient)	0.02567784838758098
Beliefs (Avatar)	-0.09561273366103458
Negative (Patient)	-0.11364269902464413
Psychotherapeutic interventions (Neutral)	0.010654286427180373
Other beliefs (Patient)	0.029167922584813755
Provocation (Avatar)	0.1866938254993158
Negation (Patient)	0.1231509841333093
Prevention (Patient)	-0.31748282158545243
Accusations (Avatar)	-0.12466383099943076
Positive (Patient)	0.4499983951894208
Questions about coping mechanisms (Avatar)	-0.35225305832560383
Counterattack (Patient)	0.10198904940934386
Questions about self-perceptions (Avatar)	0.4672076538702732
Reinforcement (Avatar)	-0.005131226285317009
Maliciousness of the voice (Patient)	-0.154875181322441

Table S2. Sample performances for each theme and class balances for each feature

Avatar theme	**Precision (VPP)**	**Recall (sensitivity)**	**F1-score (specificity)**	**Sample test size**
Accusations	0.67	0.53	0.59	30
Omnipotence	0.53	0.73	0.62	11
Beliefs	0.76	0.59	0.67	32
Active listening, empathy	0.76	0.8	0.78	20
Incitements, orders	0.67	0.91	0.77	11
Coping mechanisms	1	0.75	0.86	16
Threats	1	0.91	0.95	11
Negative emotions	0.72	0.87	0.79	15
Self-perceptions	0.65	0.65	0.65	23
Positive emotions	0.9	0.6	0.72	15
Provocation	0.43	0.71	0.54	14
Reconciliation	0.73	0.73	0.73	15
Reinforcement	0.7	0.78	0.74	18
Average scores	0.73	0.71	0.706	231
Patient themes	**Precision (VPP)**	**Recall(sensitivity)**	**F1-score (Specificity)**	**Sample test size**
Approbation	0.15	0.14	0.15	14
Self-deprecation	0.32	0.75	0.44	8
Self-appraisal	0.65	0.6	0.63	25
Other beliefs	0.62	0.58	0.6	26
Counterattack	0.5	0.62	0.56	16
Maliciousness of the voice	0.5	0.42	0.45	12
Negative	0.6	0.58	0.59	31
Negation	0.95	0.56	0.7	34
Omnipotence	0.54	0.58	0.56	12
Disappearance of the voice	0.83	0.76	0.79	25
Positive	0.71	0.88	0.79	17
Prevention	0.75	0.75	0.75	32
Reconciliation of the voice	0.55	0.75	0.63	8
Self-affirmation	0.58	0.60	0.59	25
Average scores	0.65	0.65	0.62	285

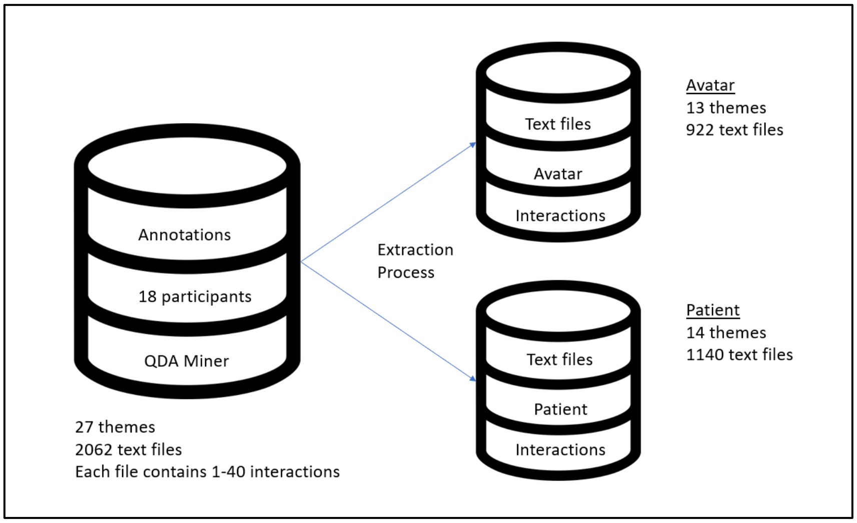

Figure 1. AT dataset and visual representation of Avatar and Patient conceptual databases.

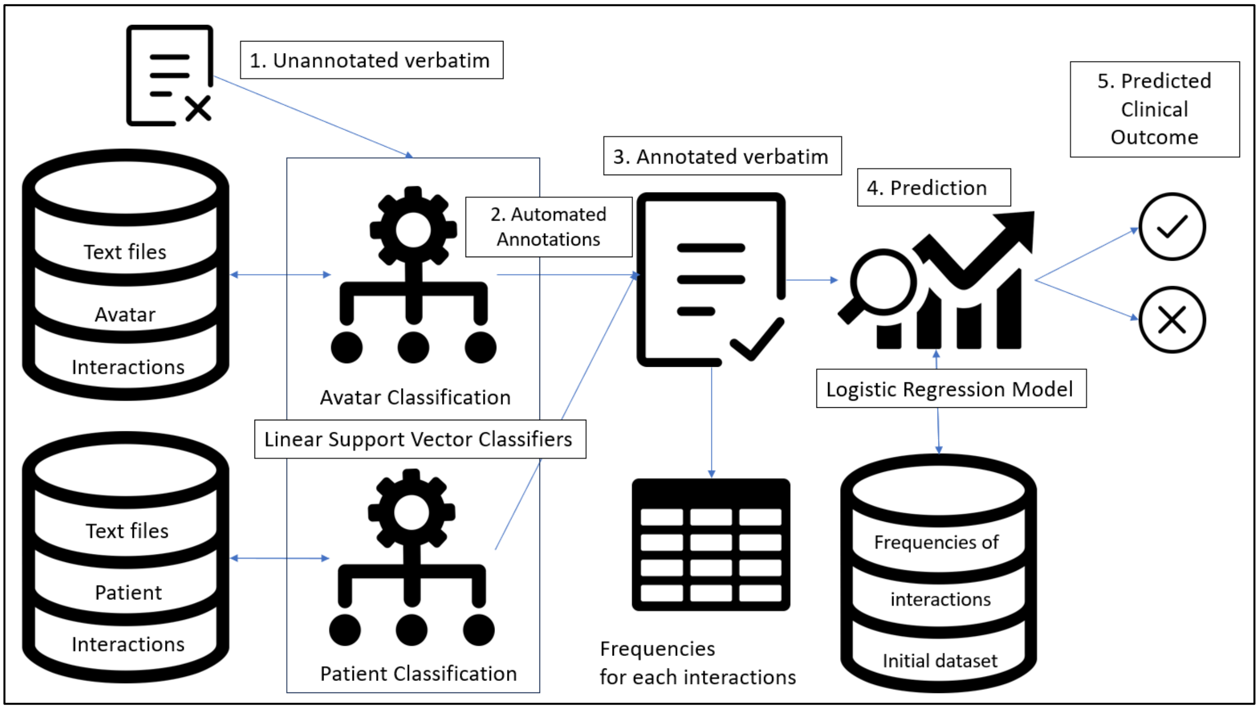

Figure 2. General flow and elements of the combination of LSVC for each conceptual database, combined to a regression algorithm.

Article 9. Exploration of the role of emotional expression of treatment-resistant schizophrenia patients having followed virtual reality therapy: a content analysis

Alexandre Hudon

Veronica Iammatteo

Sophie Rodrigues-Coutlée

Laura Dellazizzo

Sabrina Giguère

Kingsada Phraxayavong

Stéphane Potvin

Alexandre Dumais

Publié dans *BMC Psychiatry (2023).*
Article révisé par les pairs.

Abstract

Background

Emotional responses are an important component of psychotherapeutic processes. Avatar therapy (AT) is a virtual reality-based therapy currently being developed and studied for patients suffering from treatment resistant schizophrenia. Considering the importance of identifying emotions in therapeutical processes and their impact on the therapeutic outcome, an exploration of such emotions is needed.

Methods

The aim of this study is to identify the underlying emotions at the core of the patient-Avatar interaction during AT by content analysis of immersive sessions transcripts and audio recordings. A content analysis of AT transcripts and audio recordings using iterative categorization was conducted for 16 patients suffering from TRS who underwent AT between 2017 and 2022 (128 transcripts and 128 audio recordings). An iterative categorization technique was conducted to identify the different emotions expressed by the patient and the Avatar during the immersive sessions.

Results

The following emotions were identified in this study: Anger, Contempt/ Disgust, Fear, Sadness, Shame/ Embarrassment, Interest, Surprise, Joy and Neutral. Patients expressed mostly neutral, joy and anger emotions whereas the Avatar expressed predominantly interest, disgust/contempt, and neutral emotions.

Conclusions

This study portrays a first qualitative insight on the emotions that are expressed in AT and serves as a steppingstone for further investigation in the role of emotions in the therapeutic outcomes of AT.

Introduction

Schizophrenia is a chronic and complex mental disorder [1]. Despite its low prevalence of around 1% in the general population, it accounts for an annual societal cost of more than the annual cost of all cancers combined and the societal financial cost for care is directly linked to the severity of the disease [2,3,4]. This psychiatric illness is characterized by the presence of positive symptoms (i.e.: delusions, hallucinations) and negative symptoms (i.e.: alogia, avolition, blunted affect, asociality and anhedonia) [5, 6]. Positive symptoms of schizophrenia are hypothesized to be linked to an increased subcortical release of dopamine, especially in the mesolimbic region (i.e. cortical pathway involving the nucleus accumbens) [7,8,9]. This results in an increased activity of the dopaminergic receptors D2 and manifests as hallucinations and delusions [10]. It is hypothesized that the negative symptoms can either be intrinsic to the pathophysiology of schizophrenia or can be secondary symptoms that are related to various factors such as adverse effects of treatment, the environment and comorbidities [11]. Functional neuroimaging studies also support the evidence of fronto-temporal dysconnectivity in patients suffering from schizophrenia with several frontal lobe and temporal lobe abnormalities that could yield explanations for positive and negative symptoms [12, 13].

Various antipsychotic pharmaceutical approaches for positive symptoms such as hallucinations are available as first line of treatment [14, 15]. Anti-dopaminergic medication such as dopamine receptor antagonists (i.e. Risperidone, Quetiapine) and partial dopamine receptor agonists (i.e. Aripiprazole, Brexipiprazole) can be used [16]. However, around 30% of patients suffering from schizophrenia are said to be treatment resistant as they either fail to respond or only partially respond to two or more antipsychotic medications [17]. These patients tend to have poorer premorbid social functioning and represent a greater societal financial burden [18]. For these patients, Clozapine is currently the next line recommended pharmaceutical approach, but up to 60% of the patients on this medication will not respond favourably to treatment [19,20,21]. For these reasons, various adjunct approaches such as psychological therapies have been developed across the years. The main psychological intervention used for patients with treatment resistant schizophrenia is psychosis oriented cognitive-behavioral therapy (CBT) [22]. While CBT has been proven effective for reducing positive symptoms for these patients, the results remain sub-optimal and other strategies have been developed to address this limitation [23, 24].

Amongst these other strategies are virtual reality-based therapies (VRT) such as Avatar Therapy (AT). Developed by Julian Leff and his team in 2008, this psychotherapeutic approach involves the use of an immersive virtual reality system in which patients suffering from treatment resistant schizophrenia (TRS) interact with the Avatar, a virtual representation of their main persistent auditory verbal hallucination which is controlled and animated by the therapist [25]. Several studies are reporting the effectiveness of AT in the reduction of auditory and verbal hallucinations [26,27,28]. At the Institut universitaire en santé mentale de Montréal (IUSMM), AT is a protocolized therapy which is currently being studied with an undergoing trial to compare its effectiveness to CBT. It is designed as a therapeutic process that includes nine therapeutic sessions. The patients attend one AT session per week until completion of the sessions. In the first session the Avatar is being created by the therapist in collaboration with the patient, using a 3D software, to best represent their own representation of their most distressing verbal hallucination. A broad array of features can be employed (gender, facial characteristics, width, height) to design the Avatar. In the remaining eight sessions, the patients will meet and interact with the Avatar using a virtual reality headset. The Avatar is animated by the therapist and the voice of the therapist is modulated using an external voice modifier system to best represent the verbal hallucinations heard by the patient. Facial expressions of the Avatar can be modified in real-time by the therapist by using programmed dimers to modify facial features. The Avatar (including its voice) is therefore personalized for every patient.

While this technique is still being studied and developed, qualitative explorations of the therapeutic processes have been conducted to better understand the intrinsic processes linked to the improvements of patients suffering from TRS undergoing AT. Several themes related to the exchanges between the patient and the Avatar have been elicited and described, as well as the ability to automatically and adequately classify interactions such as self perceptions, beliefs about the voices and emotional responses to them [29, 30]. However, for the latter, little is known in current scientific literature regarding the expression of emotions by the patient and by the Avatar during the immersive AT sessions.

Emotional expression is crucial to the therapeutic process as it enables empathic abilities [31]. In addition to emotion attention and clarity, the integration of emotion regulation training to various CBT approaches has been associated with improvement of psychiatric and medical conditions such as persistent physical symptoms and social anxiety [32, 33].

Despite the blunted affect often portrayed by patients suffering from schizophrenia, they do experience a wide range of emotions; however, clinical access and assessment by the therapist represents a challenging and limiting factor [34]. Acoustic and vocal cues can be useful tools in the evaluation of expressed emotions [37]. Vocal cues and variation in audio samples have been studied and employed in the detection of caricatural emotions to assess them in patients with reduced affects or in patients coming from various cultural backgrounds where emotions can be expressed differently [34, 35].

Considering the importance of identifying emotions in therapeutical processes and their impact on the therapeutic outcome, an exploration of such emotions is needed. The understanding of patient’s emotions as well as the ones expressed by the Avatar in AT to further comprehend the underlying intrinsic therapeutic processes could benefit the outcome of the therapy and ultimately the patient. The aim of this study is to identify the underlying emotions at the core of the patient-Avatar interaction during AT by human-driven qualitative content analysis of immersive sessions transcripts and audio recordings. It is hypothesized that various emotions are experienced throughout the therapeutic process and that those experienced by the patients are often different than those expressed by the Avatar. To our knowledge, no study has yet explored the aspect and dynamic of emotions during AT.

Methods

Participants and sampling

Participant data used in this study originates from two completed pilot trials at the Centre de recherche de l'Institut universitaire en santé mentale de Montréal (CR-IUSMM) and one ongoing trial comparing AT to CBT [29, 36]. The data from sixteen randomly selected participants belonging to the clinical trials registered on Clinicaltrials.gov (identifier number: NCT03585127 and NCT04054778) were used. The participants included in this study were all patients selected based on the same inclusion and exclusions criteria. The inclusion criteria for this study was that the were all patients at the IUSMM, above 18 years of age, and suffering from TRS as defined by the absence of response to two or more dopaminergic antagonists. Furthermore, they all received AT between 2017 and 2022. Each patient participated in 9 psychotherapeutic sessions, each lasting one hour, of which 8 were immersive sessions in which they actively interacted with a virtual representation (the Avatar) of their auditory verbal hallucinations. The first therapeutic session is dedicated to the creation of the Avatar and was not included in the present analysis considering it does not content an immersion component. This represented a total of 128 audio recordings and transcripts. The study has been approved by the ethics committee of CR-IUSMM as part of the protocol for AT.

Data collection

All of the immersive sessions were recorded and transcribed (audio file and transcripts) by research auxiliaries. Transcripts were then counter-verified by Alexandre Hudon (AH) as per the audio recordings to ensure integrity. The auxiliaries were given a coding guideline which included specific rules to preserve the nature of the therapeutic sessions. Several elements of the coding guidelines are found in Mergenthaler and Stinson 's Psychotherapy Transcription Standards (1992) [37]. As part of the transcription rules followed by the auxiliaries, verbal utterances and paraverbal utterances were transcribed. Punctuation markers were employed to discriminate between completion of a thought and broken thoughts. Formal and structural aspects included a clear transcript heading, speaker codes and capitalization. Time recordings for each individual interactions and pauses were not part of the coding guidelines.

Data analysis

A content analysis technique using iterative categorization was conducted, as explained subsequently, to identify the different emotions expressed by the patient and the Avatar during the immersive sessions until saturation of the data [38]. Sophie Rodrigues-Coutlée (SCR), Veronica Iammateo (VI) and AH listened to the audio recordings while reading the transcript interactions to identify emotions as per the content, verbal and audio cues defined in Table 1. These emotions were selected as per Paul Ekman's and Caroll Izard's emotion-based theories [39, 40]. Turn of speech was used as the scoring unit.

-- Please insert Table 1 here --

The first step conducted in the analysis was to annotate transcripts from AT sessions with the goal of associating the emotions expressed by the patient and the Avatar (as per Table 1) to their dialogue and interactions. An initial round of annotation was done on transcripts coming from 2 randomly selected patients amongst the participants (16 transcripts and 16 audio recordings). This was done using Qualitative Data Analysis Miner software [45]. The annotations conducted by VI, SCR and AH were compared by assessing the interrater agreement of the coding of the transcripts. A Scott's Pi was employed to determine the consensus of the coding conducted by the coders [46]. The list of emotions was restructured, and the categories were updated in relation to the difference and observations found across the coders. This iterative process was repeated until the Scott's Pi obtained was deemed acceptable and data saturation was achieved. Acceptability was defined as per the SAGE Research Methods: a Scott's Pi of 0.81–1.00 is indicative of an almost perfect agreement, 0.61–0.80 of a substantial agreement, 0.41–0.60 of a moderate agreement, 0.21–0.40 of a fair agreement, 0.0–0.20 of a slight agreement and less than 0 as a poor agreement [46]. The first iteration of annotations yielded a Scott's Pi of 0.48, with the main difficulty being distinguishing between shame and embarrassment from audio recordings. Thus, these two categories were merged. The second iteration yielded a Scott's Pi of 0.51; disgust and contempt were then merged as part of this iteration. Finally, the third iteration yielded a Scott's Pi of 0.72. A total of 48 transcripts were thus annotated as part of this analysis (3 iterations). Following the third iteration, transcripts for 16 participants were annotated by AH using the final coding grid.

Results

Sample characteristics

A total of 128 transcripts and audio recordings were analyzed from 16 participants that had received AT. The sociodemographic characteristics of these participants can be found in Table 2.

-- Please insert Table 2 here --

Nine emotions were identified across the transcripts. A final description of these emotions can be found in Table 3.

-- Please insert Table 3 here --

Emotions

Anger

Anger was identified in the verbalizations and responses of most of the patients and almost never expressed by the Avatar. Anger was mostly represented by an increase in patient tone and associated with an attempt to preserve their dignity. The anger was principally aimed towards the Avatar. An example of a patient expressing anger towards the Avatar is as follows:

Patient 018: “I’m fed up with our conversations. I’d like you to leave, forever and that you stop threatening me everyday. Please leave now.”

Another source of anger is linked to the patient’s emotional attachment to the Avatar.

Patient 2020: “This is exactly why I do not like you and will never love you. It has been 20 years: you never showed any empathy towards me.”

The Avatar expresses anger solely when there is a clear provocation linked to its existence.

Avatar 2020: “But a small while ago you said I was an illness, and now you say that I do not exist? I do not follow!”

Or another clear example is in transcript 001 – T3:

Patient 001: “I hate you, go die!”

Avatar 001: “No, I hate you more! You go die!”

Contempt and disgust

Contempt and disgust are both emotions that were not equally expressed by patients and Avatar; more specifically, the patients rarely demonstrated these emotions across the immersive sessions whereas the Avatar consistently expressed them.

Patients expressed contempt and disgust mainly when confronted to a statement made by the Avatar that goes against their values or their views on a particular subject.

Example from transcript 041 – T6 :

Avatar 041: “I would love to spend much more time with you.”

Patient 041: “I do not feel that way at all.”

As per the Avatar, attempts to elicit reactions from the patients appear to be what drives the expression of contempt and disgust.

Avatar 006: “I feel it deep inside you that you have no self-esteem, which is why I can stay as long as I want in your head.”

Fear

Fear has been expressed by both the patients and the Avatar throughout the transcripts in a consistent fashion. Patients’ fear was mostly characterized by the difficulty in completing their sentences whereas the Avatar expresses fear in an exaggerated manner to empower the patient.

The participants manifested fear principally when interferences with primary needs were brought up or threatened during conversation.

Example from 039 – T7:

Avatar 039: “I can’t wait for you to be in the streets.”

Patient 039: “In the street I will not be able to have my medication because I will not have a fixed address.”

Sadness

Sadness was experienced summarily by the patients and the Avatar towards the end of the AT. In both cases, it was either in relation to belittlement, patient-Avatar affiliation or when the patient verbalized negatively valanced thoughts and claims concerning the Avatar.

Example in transcript 1041 – T8.

Avatar 001: “How come you don’t like it that much?”

Patient 001: “I don’t like my workplace, I don’t like my colleagues.”

Separation elicited by the patient towards the Avatar also yielded an expression of sadness from the Avatar.

An example can be found in transcript 001-T8:

Patient 001: “You will not be in my life anymore; you will be in a calm space.”

Avatar 001: “But, but I want to be with you! It’s been 40 years.”

Shame and embarrassment

The expression of shame and embarrassment was very rarely seen by the Avatar as compared to the patients. Themes and ideas of de-valorisation (negative perception of self) were common amongst all the patients’ transcripts. They were mostly consequent to the use of provocation or belittlement from the Avatar. As per the Avatar, the expression of shame was solely seen when the patient used self-empowerment to respond to the mere existence of the Avatar:

Patient 001: “As of now, I will not fear the words coming out of you. You are not reliable.”

Avatar 001: “I feel very small.”

Interest

Throughout the transcripts, it can be observed that the Avatar expresses interest towards the patients whereas the patients do not seem very interested in asking the Avatar open-ended questions.

During AT sessions, the Avatar attempts to elicit responses from the patients. One strategy often employed is the use of open-ended questions, which favours access to the patient’s point of view on a specific subject.

As an example, in transcript 2020 – T5:

Avatar 2020: “And why do you think we could not live together?”

Patient 2020: “Because I try to do what my doctors told me to do.”

Interest is expressed by the patients mainly when they try to obtain the point of view of the Avatar in connection with an action conducted by their auditory hallucinations.

Example from transcript 1039 – T4:

Patient 1039: “I’d like to know why you come to visit me in the evening.”

Avatar 1039: “I visit you to make fun of you.”

Surprise

The emotion of surprise was anecdotical as per patients’ expressions whereas it was a common expression identified from the Avatar. Though rare, this emotion was identified in patient’s responses when they were challenging the Avatar and the Avatar responded with a validating or positively valanced open-ended question.

An example of such interaction is identified in transcript 006-T4:

Patient 006: ‘‘Well, you are not speaking to me since a few days.”

Avatar 006: “How did you succeed in making me the quiet one?’‘.

Patient 006: “how did I succeed?”

The Avatar portrayed more exaggerated surprised responses when challenged by the patient on various topics related to their social functioning or their hallucinations.

This is well illustrated in transcript 018 – T4:

Patient 018: “The voices they come from my illness.”

Avatar 018: “What illness?”

Patient 018: “Schizophrenia.”

Avatar 018: “Are you saying that I am a disease!!?”

Joy

Joy was almost never seen as an expressed emotions from the Avatar throughout the transcripts whereas it was one of the most common emotions from the patients, especially towards the end of immersive sessions.

Patient’s expressed joy when they appeared to be in control in relation to the provocation conducted by the Avatar.

As an example, in transcript 1041 – T7:

Avatar 1041: “I told you that you would find it hard without me.”

Patient 1041: “Oh but no worries, I will be fine! Time will act against you.”

As for the Avatar, joy was expressed solely to try and elicit further reaction from the patient when faced with a closed-ended affirmation.

In transcript 018- T4 demonstrates such interaction:

Avatar 018: “Me, I love to have power.”

Patient 018: “I have no doubt about this.”

Avatar 018: “And I’m so happy because you gave me this power.”

Neutral

Neutral verbalizations were the most popular ones observed amongst the patients. Normal tone with no particular associated content constitute the type of interactions that were encountered, especially at the beginning of the immersive sessions.

These interactions where also linked to the technical aspects of the immersive sessions in which the Avatar might validate certain specificities such as display brightness.

For example, in transcript 1039- T5:

Avatar 1039: “Do you see me?”

Patient 1039: “Yes. I see you.”

Discussion

The objective of this study was to explore the emotions of patients' suffering of TRS and that have undergone AT. It was also designed to identify the emotions expressed by the Avatar throughout the immersive sessions. Nine emotions were identified across the transcripts: Anger, Contempt/ Disgust, Fear, Sadness, Shame/ Embarrassment, Interest, Surprise, Joy and Neutral. Neutral, joy and anger were the emotions that were mostly expressed by the patients. As for the Avatar, expression of interest, disgust/contempt and neutral were amongst the emotions the most annotated across the transcripts.

Patients and Avatars in this study expressed various emotions during the psychotherapeutic process of AT. A recent thematic qualitative evaluation of AT involving views of 15 patients on the therapeutic process identified voice embodiment and associated emotions as a major theme, considering that the voice of the Avatar triggers emotional responses [47]. This can explain why various emotional responses were identified in the presented study and why there seems to be different links between specific verbalizations by the Avatar and the emotional response expressed by the patient. Another recent study explored emotion elicitation in virtual reality for 11 participants and demonstrated the possibility to elicit fear and anger in a secured immersive environment [48]. Similarly, virtual environments themselves have been found to intrinsically elicit emotions across patients in virtual reality settings [49].

While similar studies are not identified for patients suffering from TRS, the polarity between anger and joy, the most frequently identified emotions across the transcripts for the patients (after neutral emotions), might be explained by the neurophysiological changes observed in patients suffering from TRS. Neuroimaging studies have reported that emotionally laden images elicited hyper-activation in the dorso-medial prefrontal context and left cerebellum in TRS patients [50]. In another study, weaker cerebellum activity presented with deficits in emotion recognitions in schizophrenia [51]. Since the Avatar is visually represented and presents modifications of facial expressions, this might trigger these hyper-activations and oscillations between emotional responses of anger, neutral versus joy, rarely including the other emotions. It is also important to note that emotion identification is possible in patients suffering from schizophrenia and although their affect might not display their emotional processes and responses, they can feel them [52]. As for the fear response in the patients, it has been found that virtual avatar and human responses can both elicit the same response in the amygdala even if the avatar are overly anthropomorphic [53]. Further understanding of these emotional responses could be elicited using a wider array of parameters such as heart rate, body temperature, gesture, and overall behavior.

As for the Avatar, as part of the therapeutical processes, it is important for the therapist to elicit patient reactions. Interest, similarly, to most positive emotions, is an emotion that can be strategically employed to create the therapeutic alliance and to reduce the anxiety and fear linked to the therapeutic sessions themselves [54, 55]. It is also an emotion that can help the therapist to use real and personalized examples to confront or validate the patient, which may explain why interest is such a frequently annotated emotion throughout Avatar transcripts. The emotional expressiveness is an important component of the therapy and it has been demonstrated that simulation of Avatars through virtual reality is a way to train patients suffering from schizophrenia in their abilities to recognize certain emotions [56].

Limitations

This is an exploratory study and the lack of generalisation of the results can limit its interpretation. Considering the emotional responses, the emotion identification process could have been biased by the coders' own understanding and perception of emotions. This has been mitigated using an emotional grid that was defined and the interrater analysis. This study does not include the patient's own labelling and identification of his or her emotions underlying their reactions and transcripts and there could thus be a mismatch between the coders' perceptions and patients' perceptions of their own emotions. Another limitation is that visual cues were not taken into consideration which limits the analysis to the content of the verbatims and the audio transcripts. The random selection of the participants could yield to bias considering that emotionally tinged speech can vary for men and women [57].

Conclusion

To conclude, the main objective of this study was to explore the emotions of TRS patients and the Avatar in AT. The use of an iterative categorization content analysis process enabled the identification of nine emotions. The nine emotions were identified across the transcripts and appeared to be linked with Avatar-Patient dynamics as particular emotional responses were often seen in contexts of provocation or belittlement. Emotional expressions were more polarized in the patients where anger, joy and neutrality were predominant as compared to the Avatar where interest, disgust/contempt and neutrality were more observed. While this study portrays a first qualitative insight on the emotions that are expressed in AT, further studies are needed to assess their role in the treatment response associated to the immersive session of AT.

Data Availability

The datasets generated and/or analysed during the current study are not publicly available due to patients' privacy but are available from the corresponding author on reasonable request.

Abbreviations

AT: Avatar Therapy
CBT: Cognitive-behavioral therapy
TRS: Treatment resistant schizophrenia

Funding

This work was indirectly supported by Le Fonds de recherche du Québec – Santé (FRQS); Otsuka Canada Pharmaceutical Inc.; Chaire Eli Lilly Canada de recherche en schizophrénie; Fondation Pinel; Services et recherches psychiatriques AD; Fonds d'excellence en recherche Apogée Canada. These fundings bodies had no part in the data collection, analysis, interpretation of data and in writing the manuscript.

Contributions

All authors took part in the conceptual design of this study. AH, VI and SRC performed the data collection. AH, VI, SRC, LD, SG, SP and AD participated in the data analysis. AH wrote the manuscript. All the authors participated in the revision of the manuscript. KP conducted the administrative tasks relevant to this study. SP and AD supervised this study. All authors read and approved the final manuscript.

Ethics declarations

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

This study was approved by the institutional ethical committee. Patients that are part of this study were selected based on the proof-of-concept trial from Percy du Sert 2018 's study and Dellazizzo 2021s study [28, 39]. The trial was conducted in accordance with the Declaration of Helsinki and was approved by the institutional ethical committee (CER IPPM 16-17-06). The study has been approved by the ethics committee of CR-IUSMM. We obtained written informed consent from all patients.

Consent for publication

Not applicable.

References

1. Kahn RS, Sommer IE, Murray RM, Meyer-Lindenberg A, Weinberger DR, Cannon TD, et al. Schizophrenia. Nat Rev Dis Primers. 2015;1:15067.

2. Jin H, Mosweu I. The Societal Cost of Schizophrenia: A Systematic Review. Pharmacoeconomics. 2017;35(1):25-42.

3. Kadakia A, Catillon M, Fan Q, Williams GR, Marden JR, Anderson A, et al. The Economic Burden of Schizophrenia in the United States. J Clin Psychiatry. 2022;83(6).

4. Thaker GK, Carpenter WT, Jr. Advances in schizophrenia. Nat Med. 2001;7(6):667-71.

5. McCutcheon RA, Reis Marques T, Howes OD. Schizophrenia-An Overview. JAMA Psychiatry. 2020;77(2):201-10.

6. Abplanalp SJ, Braff DL, Light GA, Nuechterlein KH, Green MF, Consortium on the Genetics of S. Understanding Connections and Boundaries Between Positive Symptoms, Negative Symptoms, and Role Functioning Among Individuals With Schizophrenia: A Network Psychometric Approach. JAMA Psychiatry. 2022;79(10):1014-22.

7. Brisch R, Saniotis A, Wolf R, Bielau H, Bernstein HG, Steiner J, et al. The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. Front Psychiatry. 2014;5:47.

8. McCollum LA, Roberts RC. Uncovering the role of the nucleus accumbens in schizophrenia: A postmortem analysis of tyrosine hydroxylase and vesicular glutamate transporters. Schizophr Res. 2015;169(1-3):369-73.

9. Danielsson K, Stomberg R, Adermark L, Ericson M, Soderpalm B. Differential dopamine release by psychosis-generating and non-psychosis-generating addictive substances in the nucleus accumbens and dorsomedial striatum. Transl Psychiatry. 2021;11(1):472.

10. McCutcheon RA, Abi-Dargham A, Howes OD. Schizophrenia, Dopamine and the Striatum: From Biology to Symptoms. Trends Neurosci. 2019;42(3):205-20.

11. Correll CU, Schooler NR. Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. Neuropsychiatr Dis Treat. 2020;16:519-34.

12. Butler T, Weisholtz D, Isenberg N, Harding E, Epstein J, Stern E, et al. Neuroimaging of frontal-limbic dysfunction in schizophrenia and epilepsy-related psychosis: toward a convergent

neurobiology. Epilepsy Behav. 2012;23(2):113-22.

13. John JP. Fronto-temporal dysfunction in schizophrenia: A selective review. Indian J Psychiatry. 2009;51(3):180-90.

14. Smith RC, Leucht S, Davis JM. Maximizing response to first-line antipsychotics in schizophrenia: a review focused on finding from meta-analysis. Psychopharmacology (Berl). 2019;236(2):545-59.

15. Zhu Y, Krause M, Huhn M, Rothe P, Schneider-Thoma J, Chaimani A, et al. Antipsychotic drugs for the acute treatment of patients with a first episode of schizophrenia: a systematic review with pairwise and network meta-analyses. Lancet Psychiatry. 2017;4(9):694-705.

16. Lally J, MacCabe JH. Antipsychotic medication in schizophrenia: a review. Br Med Bull. 2015;114(1):169-79.

17. Pandey A, Kalita KN. Treatment-resistant schizophrenia: How far have we traveled? Front Psychiatry. 2022;13:994425.

18. Nucifora FC, Jr., Woznica E, Lee BJ, Cascella N, Sawa A. Treatment resistant schizophrenia: Clinical, biological, and therapeutic perspectives. Neurobiol Dis. 2019;131:104257.

19. Bioque M, Parellada E, Garcia-Rizo C, Amoretti S, Fortea A, Oriolo G, et al. Clozapine and paliperidone palmitate antipsychotic combination in treatment-resistant schizophrenia and other psychotic disorders: A retrospective 6-month mirror-image study. Eur Psychiatry. 2020;63(1):e71.

20. Correll CU, Rubio JM, Inczedy-Farkas G, Birnbaum ML, Kane JM, Leucht S. Efficacy of 42 Pharmacologic Cotreatment Strategies Added to Antipsychotic Monotherapy in Schizophrenia: Systematic Overview and Quality Appraisal of the Meta-analytic Evidence. JAMA Psychiatry. 2017;74(7):675-84.

21. Potkin SG, Kane JM, Correll CU, Lindenmayer JP, Agid O, Marder SR, et al. The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. NPJ Schizophr. 2020;6(1):1.

22. Bighelli I, Salanti G, Huhn M, Schneider-Thoma J, Krause M, Reitmeir C, et al. Psychological interventions to reduce positive symptoms in schizophrenia: systematic review and network meta-analysis. World Psychiatry. 2018;17(3):316-29.

23. Bighelli I, Huhn M, Schneider-Thoma J, Krause M, Reitmeir C, Wallis S, et al. Response rates

in patients with schizophrenia and positive symptoms receiving cognitive behavioural therapy: a systematic review and single-group meta-analysis. BMC Psychiatry. 2018;18(1):380.

24. Jauhar S, McKenna PJ, Radua J, Fung E, Salvador R, Laws KR. Cognitive-behavioural therapy for the symptoms of schizophrenia: systematic review and meta-analysis with examination of potential bias. Br J Psychiatry. 2014;204(1):20-9.

25. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff AP. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? Psychosis. 2014;6(2):166-76.

26. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry. 2018;5(1):31-40.

27. Aali G, Kariotis T, Shokraneh F. Avatar Therapy for people with schizophrenia or related disorders. Cochrane Database Syst Rev. 2020;5(5):CD011898.

28. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophr Res. 2018;197:176-81.

29. Beaudoin M, Potvin S, Machalani A, Dellazizzo L, Bourguignon L, Phraxayavong K, et al. The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clin Psychol Psychother. 2021;28(3):500-18.

30. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach. Health Informatics J. 2022;28(4):14604582221142442.

31. Abargil M, Tishby O. How therapists' emotion recognition relates to therapy process and outcome. Clin Psychol Psychother. 2022;29(3):1001-19.

32. Kleinstauber M, Allwang C, Bailer J, Berking M, Brunahl C, Erkic M, et al. Cognitive Behaviour Therapy Complemented with Emotion Regulation Training for Patients with Persistent Physical Symptoms: A Randomised Clinical Trial. Psychother Psychosom. 2019;88(5):287-99.

33. Butler RM, Boden MT, Olino TM, Morrison AS, Goldin PR, Gross JJ, et al. Emotional clarity and attention to emotions in cognitive behavioral group therapy and mindfulness-based stress

reduction for social anxiety disorder. J Anxiety Disord. 2018;55:31-8.

34. Whiting CM, Kotz SA, Gross J, Giordano BL, Belin P. The perception of caricatured emotion in voice. Cognition. 2020;200:104249.

35. Kikutani M, Ikemoto M. Detecting emotion in speech expressing incongruent emotional cues through voice and content: investigation on dominant modality and language. Cogn Emot. 2022;36(3):492-511.

36. Dellazizzo L, Potvin S, Phraxayavong K, Dumais A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. NPJ Schizophr. 2021;7(1):9.

37. Mergenthaler E, Stinson C. Psychotherapy Transcription Standards. Psychotherapy Research. 2010;2(2):125-42.

38. Neale J. Iterative categorization (IC): a systematic technique for analysing qualitative data. Addiction. 2016;111(6):1096-106.

39. Ekman P. Facial expression and emotion. Am Psychol. 1993;48(4):384-92.

40. Izard CE. Emotion theory and research: highlights, unanswered questions, and emerging issues. Annu Rev Psychol. 2009;60:1-25.

41. Coan JA, Gottman JM. The Specific Affect Coding System (SPAFF). Handbook of emotion elicitation and assessment. Series in affective science. New York, NY, US: Oxford University Press; 2007. p. 267-85.

42. Sbattella L, Colombo L, Rinaldi C, Tedesco R, Matteucci M, Trivilini A, editors. Extracting Emotions and Communication Styles from Vocal Signals. PhyCS; 2014.

43. Rochman D, Amir O. Examining in-session expressions of emotions with speech/vocal acoustic measures: an introductory guide. Psychother Res. 2013;23(4):381-93.

44. Lyons M, Aksayli ND, Brewer G. Mental distress and language use: Linguistic analysis of discussion forum posts. Computers in Human Behavior. 2018;87:207-11.

45. Chomczynski P. QDA MINER – The Mixed Method Solution for Qualitative Analysis by Provalis Research. Qualitative Sociology Review. 2008;4(2):126-9.

46. Given LM. The Sage encyclopedia of qualitative research methods: Sage publications; 2008.

47. Rus-Calafell M, Ehrbar N, Ward T, Edwards C, Huckvale M, Walke J, et al. Participants' experiences of AVATAR therapy for distressing voices: a thematic qualitative evaluation. BMC Psychiatry. 2022;22(1):356.

48. Susindar S, Sadeghi M, Huntington L, Singer A, Ferris TK. The Feeling is Real: Emotion Elicitation in Virtual Reality. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2019;63(1):252-6.

49. Felnhofer A, Kothgassner OD, Schmidt M, Heinzle A-K, Beutl L, Hlavacs H, et al. Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. International Journal of Human-Computer Studies. 2015;82:48-56.

50. Potvin S, Tikasz A, Lungu O, Dumais A, Stip E, Mendrek A. Emotion processing in treatment-resistant schizophrenia patients treated with clozapine: An fMRI study. Schizophr Res. 2015;168(1-2):377-80.

51. Mothersill O, Knee-Zaska C, Donohoe G. Emotion and Theory of Mind in Schizophrenia-Investigating the Role of the Cerebellum. Cerebellum. 2016;15(3):357-68.

52. Gica S, Poyraz BC, Gulec H. Are emotion recognition deficits in patients with schizophrenia states or traits? A 6-month follow-up study. Indian J Psychiatry. 2019;61(1):45-52.

53. Kegel LC, Brugger P, Fruhholz S, Grunwald T, Hilfiker P, Kohnen O, et al. Dynamic human and avatar facial expressions elicit differential brain responses. Soc Cogn Affect Neurosci. 2020;15(3):303-17.

54. Prusinski T. The Strength of Alliance in Individual Psychotherapy and Patient's Wellbeing: The Relationships of the Therapeutic Alliance to Psychological Wellbeing, Satisfaction With Life, and Flourishing in Adult Patients Attending Individual Psychotherapy. Front Psychiatry. 2022;13:827321.

55. Ardito RB, Rabellino D. Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. Front Psychol. 2011;2:270.

56. Souto T, Silva H, Leite A, Baptista A, Queirós C, Marques A. Facial Emotion Recognition: Virtual Reality Program for Facial Emotion Recognition—A Trial Program Targeted at Individuals With Schizophrenia. Rehabilitation Counseling Bulletin. 2019;63(2):79-90.

57. Kraemer S, Lihl M, Mergenthaler E. Schlüsselstunden im Verlauf kognitiver

Verhaltenstherapie von schizophrenen Patienten: Ein Beitrag zur Prozessforschung. Verhaltenstherapie. 2007;17(2):90-9.

Figures and Tables

Table 1. Emotions classification as per content, verbal and audio cues.

Emotions	**Cues**
Anger	Irregular or fast speech with an attempt to stop individual violation or preserve dignity(41, 42).
Contempt	Humiliation, attempt to belittle or achieve superiority. Often with lack of respect or cruelty. Sarcasm, provocation, one-sided humor and hostility often seen(41).
Disgust	Involuntary repulsion(41).
Fear	Superficial and rapid breathing, accelerated speech, incomplete declarations, nervous laugher, difficulty to express themselves(41, 42).
Sadness	Slower or irregular speech, monotone, increases in pauses or duration of syllable pronunciation. Tendency to use a larger number of possessive pronouns. (42, 43, 44).
Shame	Feeling of uselessness, powerlessness, negative self-evaluation(41).
Embarrassment	Inhibition of the capacity to do or say their thoughts, insecurity(41).
Interest	Open-ended questions, validations, respectful engagement in the discussion(41).
Surprise	Audible panting, brief emotion, fast and spontaneous(41).
Joy	Augmentation of the speed pressure, satisfaction, exclamation, excitement, laugher(42).
Neutral	Interactions with no speech or content valence(41). Everything that cannot be identified as one of the above emotions are included in this category.

Table 2. Summary of participants' characteristics

Characteristics	Value (N=16)
Sex (male, female)	13,3
Age (mean in years)	40.8 ± 9.4
Education (mean in years)	12.8 ± 3.0
Ethnicity (Caucasian, others)	93.8%,6.2%
% on Clozapine	37.5%

Table 3. Redefined coding grid as per the iterative process.

Emotions	Cues
Anger	The interlocutor presented an irregular or a faster response than usual with a clear attempt to preserve their dignity or cease an individual violation of their integrity.
Contempt/ Disgust	The interlocutor uses humiliation, belittlement or lack respect to their interlocutor. They may use sarcasm, provocation, hostility, or sound repulsed.
Fear	The interlocutor demonstrates difficulty in completing declarations, laugh nervously, portray difficulty to express themselves.
Sadness	The interlocutor adopts a slower speech with a monotone tonality. There is a clear increase in pauses or an increase in the use of possessive singular pronouns.
Shame/ Embarrassment	The interlocutor's speech content depicts uselessness, powerlessness, and a negative perception of Self.
Interest	The interlocutor uses open-ended questions, shows interest to what their peer is saying and engages respectfully in the discussion.
Surprise	The interlocutor has a spontaneous response that displays a brief emotion, with an acceleration of speech and audible panting.
Joy	The interlocutor demonstrates a clear increase in satisfaction, excitement, laughs and has an increase in the speed of their speech.

Neutral	The interlocutor's interaction has no speech or content valence such as a change in tonality, in speed or engages conversation with neutral content.

Article 10. Dyadic Interactions of Treatment-Resistant Schizophrenia Patients Having Followed Virtual Reality Therapy: A Content Analysis

Alexandre Hudon

Jonathan Couture

Laura Dellazizzo

Mélissa Beaudoin

Kingsada Phraxayavong

Stéphane Potvin

Alexandre Dumais

Publié dans *Journal of Clinical Medicine (2023).*
Article révisé par les pairs.

Abstract

(1) Background: Very little is known about the inner therapeutic processes of psychotherapy interventions for patients suffering from treatment-resistant schizophrenia. Avatar therapy (AT) is one such modalities in which the patient is undergoing immersive sessions in which they interact with an Avatar representing their main persistent auditory verbal hallucination. The aim of this study is to identify the most prevalent dyadic interactions between the patient and the Avatar in AT for patient's suffering from TRS. (2) Methods: A content analysis of 256 verbatims originating from 32 patients who completed AT between 2017 and 2022 at the Institut universitaire en santé mentale de Montréal was conducted to identify dyadic interactions between the patients and their Avatar. (3) Results: Five key dyads were identified to occur on average more than 10 times for each participant during the immersive sessions across their AT: (Avatar: Reinforcement, Patient: Self-affirmation), (Avatar: Provocation, Patient: Self-affirmation), (Avatar: Coping mechanisms, Patient: Prevention), (Patient: Self-affirmation, Avatar: Reinforcement), and (Patient: Self-appraisal, Avatar: Reinforcement). (4) Conclusion: These dyads offer a first qualitative insight to the interpersonal dynamics and patient-avatar relationships taking place during AT. Future studies on the implication of such dyadic interactions with the therapeutic outcome of AT should be conducted considering the importance of dyadic relationships in psychotherapy.

Keywords

psychotherapy; virtual reality therapy; auditory hallucinations; schizophrenia; avatar therapy; dyads; dyadic relationship

Introduction

Psychotherapy is a complementary treatment to medication for many psychopathologies in mental health [1,2]. Individual psychotherapy is defined as an approach in which a therapist and patient are interacting together to improve psychopathologic conditions and functional impairment through the therapeutic alliance [3]. The mechanisms establishing the success of a psychotherapy are widely debated [2,4]. Despite the debates, the therapeutic bond between the therapist and the patient is perceived to be a major factor resulting in behavioral, social, or cognitive changes [5,6]. As part of this bond, the notion of transference and countertransference is relevant. Derived from psychodynamic paradigms, transference is defined as a process in which individuals displace patterns of behavior that evolve through interaction with significant figures in childhood onto other persons in their current lives. On the other hand, countertransference is known to be a corresponding response of the therapist following the transference [7,8,9]. These notions have clinical implications as they provide insight to the inner world of all parties involved in the therapeutic process [10,11]. For example, regarding transference in classical cognitive-behavioral therapy (CBT), it has been suggested that therapists should not deliberately provoke or ignore their patient; instead, they should be aware of their own feeling and monitor them [12]. Psychotherapeutic approaches can also be used for complex mental illness as an adjunct to psychopharmaceutical recommendations. Notably, a 25-year systematic review and exploratory meta-analysis reported that CBT was the most frequently recommended psychotherapy intervention for patients suffering from treatment-resistant schizophrenia (TRS) [13].

As stated above, one example of severe and complex mental illness is TRS. Schizophrenia is part of the psychotic disorders and is characterized by the presence of positive symptoms and negative symptoms, with an occupational or social dysfunction, with continued and persistent disturbance for more than 6 months [14]. Frequently seen positive symptoms are hallucinations (mostly auditory) and delusions, both of which are hypothesized to be due to hyperactive dopaminergic activation in the mesolimbic system [15,16]. Negative symptoms consist of five constructs: blunted affect, asociality, anhedonia, alogia, and avolition [17]. While various definitions exist for TRS, one that is widely used is a documented failure to two or more antipsychotics [18]. Around 20–30% of patients suffering from schizophrenia will evolve to TRS and about 40–70% of these patients will not respond to the treatment of choice (Clozapine) [19]. Therefore, there was a crucial need to identify new treatment approaches for these patients [19,20,21].

Several studies demonstrated benefits associated with psychotherapeutic approaches for targeting the positive symptoms of patients suffering from TRS [22,23]. Such techniques include CBT designed for psychosis. However, considering the mitigated results, further techniques were developed, such as avatar therapy (AT) [24,25]. This immersive therapeutical approach was developed by Julian Leff in 2008 [26]. In AT, the patients interact with a virtual representation of the patient's most disturbing auditory verbal persistent hallucination, which we refer to as ''the Avatar''. Pilot studies investigating the effects of AT demonstrated an improvement in patients suffering from TRS by reducing their auditory hallucination with a large effect size and positive changes in their quality of life [27,28,29]. Moreover, attempts were made to further understand the intrinsic psychotherapeutic processes. A first qualitative analysis exploring the dialogue components of AT identified five major themes amongst the patients: emotional responses to the voices, beliefs about voices and schizophrenia, self-perceptions, coping mechanisms, and aspirations [30]. These results were further developed by Beaudoin and her team by conducting a content analysis of the verbatims of 18 patients who received AT. In doing so, they were able to sub-divide the previously identified themes as well as explore the themes emerging from the avatar's interactions [31]. Provocation, mainly the act of provoking the patient intentionally to elicit a reaction, was identified to be one of the most frequent interactions. However, the tuple (combination of two interactions) of interactions between the avatar and the patients have never been explored together as unique units of interactions known as dyads. The consideration of this combination is relevant in order to better understand the inner processes of AT considering that, in comparison with CBT, the therapist takes a more active role in coaching the patient as to how to respond to their auditory hallucinations. This may influence how the patient will respond to their visual representation of their auditory hallucinations. This is important, as it reports to the notion of transference and countertransference and its integration across a virtual environment. Furthermore, it has been previously observed that social interactions during psychotherapeutic interactions in virtual reality can be explored in different psychopathologies, such as in social anxiety. The engagement itself between the patient and a virtual human demonstrated a reduction in social anxiety in 18 participants suffering from high levels of social anxiety [32].

The aim of this study is to identify the most prevalent dyadic interactions between the patient and the Avatar in AT for patients suffering from TRS. It is hypothesized that certain dyadic interactions, such as the ones implying a provocation from the Avatar and a counterattack from the patient, will be predominant considering the importance of this theme in the therapy. Considering the immersive nature of this therapeutic approach, the dyadic interventions could provide additional understanding of the psychotherapeutic processes of AT to further explore the nature of transference and countertransference elements of this therapeutic approach.

Materials and Methods

Participants and Recruitment

The data used in this study were derived from participants who received AT in the context of a pilot clinical trial as well as one ongoing trial comparing AT to CBT, all conducted at the Centre de recherche de l'Institut universitaire en santé mentale de Montréal (CR-IUSMM) [28,29]. The participants all belong to the clinical trials registered on Clinicaltrials.gov (identifier numbers: NCT03585127 and NCT04054778). The delivery of AT was the same across the two trials as protocolized. Participants in these studies received nine one-hour psychotherapeutic sessions, of which eight were immersive sessions. The patients received AT from one of the only two therapists trained to provide AT at CR-IUSMM. During these sessions, they interacted with a virtual representation of their auditory verbal hallucinations: the Avatar. The participants included in this study were all patients at the IUSMM, above 18 years of age, and suffering from TRS as defined by the absence of response to two or more dopaminergic antagonists and receiving AT between 2017 and 2022. Study has been approved by the ethics committee of CR-IUSMM as part of the protocol for AT.

Data Collection

The immersive sessions of 32 patients who underwent AT were transcribed to verbatims from audio recordings by research auxiliaries. The verbatims were then verified by AH to ensure integrity of the transcriptions. This yielded 256 verbatims representing over 230 h of immersion in AT. Annotations of the interactions between the patients and the avatars were classified as per the 27 themes described in Beaudoin et al. [31]. The themes are presented in Table 1 for the Avatar and Table 2 for the patients.

-- Please insert Table 1 here --
-- Please insert Table 2 here --

The annotation of the verbatims was done automatically by using a peer-reviewed trained linear support vector classifier, previously trained and implemented on a dataset for AT using Python 3.9 with the Scitkit-Learn open library and a 10-k fold cross validation [33].

The performance for these initial automated annotations for the Avatar and the patients 'themes is presented in Table 3. The results are comparable to the ones obtained in the description of the original study on automated classification of interaction for AT [33].

-- Please insert Table 3 here --

Data Analysis

Dyadic Interactions

A dyad is the combination of two successive themes that results from the interaction of the Avatar with the patient or vice-versa. It is, therefore, the result of the engagement between the patient and the Avatar over two consecutive interactions.

For example, an (Avatar, Patient) dyad could be represented as follows:

(Avatar: Accusations, Patient: Self-affirmation).

Conversely, a (Patient, Avatar) dyad could be represented as follows:

(Patient: Prevention, Avatar: Beliefs).

Analysis of Dyads

Using a Python script developed by A.H., dyads of interactions between the patients and their Avatar were compiled into an Excel spreadsheet and the frequencies of apparition of each dyad per verbatim were counted. The mean frequencies of each dyad for each participant were computed.

Dyads of interactions between the therapists were compiled into an Excel spreadsheet using a Python script developed by A.H. and was compared manually for four patients by J.C. The frequencies of apparition of each dyad per verbatim were counted. The mean frequencies of each dyad for each participant were computed. Dyads were selected for this study if they had a mean frequency above 10. The mean frequency above 10 was selected as most of the dyads had a mean frequency of four and below in the descriptive statistical observations of the identified dyads.

Results

Sample Characteristics

From the 256 verbatims of the 32 patients who underwent AT, a total of 1117 dyads were identified. Patients' characteristics are presented in Table 4. Out of the identified dyadic interactions, only five dyads presented a mean frequency above 10. Figure 1 displays the identified dyads as well as their mean frequencies.

-- Please insert Figure 1 here --

-- Please insert Table 4 here --

Dyadic Interactions

Avatar: Reinforcement, Patient: Self-Affirmation

The dyadic interaction unit with the highest mean frequency is (Avatar: Reinforcement, Patient: Self-affirmation). It is characterized by an attempt of the Avatar to reinforce a statement made by the patient followed by a self-affirmation statement expressed by the patient. Such circumstances can be noted when the Avatar specifies an action that was completed by the patient to elicit further affirmations from the patient regarding this action.
For example, in verbatim 21005—T4:

Avatar 21005 (Reinforcement): “That’s it. You can tell me directly like this if you have something to tell me”.

Patient 21005 (Self-affirmation): “Well, it’s just that I felt badly when you elevated the tone of your voice. You seemed angry and it affected me”.

Avatar: Provocation, Patient: Self-Affirmation

A provocation expressed by the Avatar followed by a patient’s self-affirmation yields another frequently identified dyadic interaction: (Avatar: Provocation, Patient: Self-affirmation). Provocations are frequent in AT and this is often performed by the Avatar to trigger an emotional response from the patient.

An example is found in verbatim 004-T5:

Avatar 004 (Provocation): “You are just someone who is not convincing at all”.

Patient 004 (Self-affirmation): “But if you think about it, maybe I should be changing something because I recognize that it does not make sense”.

Avatar: Coping Mechanisms, Patient: Prevention

Coping mechanisms as per the Avatar's theme is often under the form of a question related to the behavior of the patient or cognitive processes regarding a particular response. The dyadic interaction (Avatar: Coping mechanisms, Patient: Prevention) identifies a question (as stated above) followed by a response from the patient that states a prevention strategy they believe will help them deal with their auditory hallucinations.

This is illustrated in verbatim 008-T4:

Avatar 008 (Coping mechanisms): "How would you want me to leave your brain?"

Patient 008 (Prevention): "I think I'll just stop listening. This way you will leave my brain".

Patient: Self-Affirmation, Avatar: Reinforcement

While the (Avatar: Reinforcement, Patient: Self-affirmation) dyadic interaction was the most prevalent in terms of its mean frequency, its reverse was also identified substantially. The (Patient: Self-affirmation, Avatar: Reinforcement) dyad implies a positive assertion from the patient that demonstrates some self-confidence, followed by a further assertion from the Avatar to encourage more positive assertions.

Such an occurrence can be identified in verbatim 1015-T5:

Patient 1015 (Self-affirmation): "I believe this is right. It changed something for me".

Avatar 1015 (Reinforcement): "You seem very decided".

Patient: Self-Appraisal, Avatar: Reinforcement

When the patient compliments themselves or agrees with a compliment made by the Avatar and this is followed by a positive assertion from the Avatar, the dyadic interaction (Patient: Self-appraisal, Avatar: Reinforcement) is observed. This is often seen when the Avatar attempts to validate a positive statement made by the patient to elicit a further assertion or emotional response.

An example is found in verbatim 1011-T4:

Patient 1011 (Self-appraisal): “I am very kind and respectful!”

Avatar 1011 (Reinforcement): “Oh! You are starting to affirm yourself!”

Discussion

The main objective of this study was to identify the dyadic interactions between the patient and the Avatar that occur the most frequently in AT. The in-depth analysis of 256 verbatims from 32 patients were automatically annotated and the different types of interactions were compiled into dyadic interactions for further analysis. Amongst the identified dyads, five dyadic interactions occurred more than 10 times on average per participant. These dyadic interactions are, in decreasing order of prevalence: (Avatar: Reinforcement, Patient: Self-affirmation), (Avatar: Provocation, Patient: Self-affirmation), (Avatar: Coping mechanisms, Patient: Prevention), (Patient: Self-affirmation, Avatar: Reinforcement), and (Patient: Self-appraisal, Avatar: Reinforcement).

Current studies of psychotherapeutic processes urge the consideration of dyadic interactions to better understand co-regulation, receptiveness, and influence in interpersonal relationships [34]. During AT, the patients and the Avatar (animated by the therapist) are members of a dyadic relationship. This relationship rather than individual interactions can influence perceptions and interpersonal dynamics during therapy. For example, the authors of a recent study, comprising 12 general practitioners and 189 patients, conducted an analysis of dyadic relations to assess interpersonal trust in doctor and patient relationships and concluded that relationship and reciprocity effects converged with perception of trust [35]. As we can see in the dyadic interaction (Avatar: Reinforcement, Patient: Self-affirmation), in which a reinforcement assertion is followed by a positive self-assertion, there is a transference of positivity across the members of the dyadic relationship. It has been demonstrated that positive transference can act as a moderator during therapy and can help the therapist in the moderation of the treatment phase and depth [36]. Therefore, by reinforcing the patient, the therapist (as the Avatar) might help the patients in acquiring confidence and learning to affirm themselves. A similar hypothesis can be applied to the reverse dyad (Patient: Self-affirmation, Avatar: Reinforcement) and the related dyad (Patient: Self-appraisal, Avatar: Reinforcement), in which the Avatar might attempt to build on this self-confidence to elicit further self-affirmation.

Similarly, provocation is part of AT in order to mimic the negative experiences of auditory hallucinations experienced as reported by the patients. Provocation is one of the key factors known to provoke anger in human relationships [37]. However, as per the (Avatar: Provocation, Patient: Self-affirmation), the patient's common reaction in AT is to respond expressing self-affirmation. Patients suffering from schizophrenia are reported to have difficulties in adequately identifying social cues [38,39]. Therefore, a provocation expressed by the Avatar could be misinterpreted as an opportunity to affirm or re-affirm a statement regarding their experience with their auditory hallucinations. In the community, patients suffering from auditory hallucinations often adopt safety-seeking behaviors, which maintains their distorted beliefs regarding their hallucinations [40]. In AT, they cannot adopt these safety-seeking behaviors and are in a secured environment to react to their auditory hallucinations, which might explain why self-affirmations are more prevalent. Another hypothesis is that in the AT protocol, the therapist encourages the patient to affirm themselves when the Avatar is talking to them. In that sense, it could indicate that the patient is responding adequately to the coaching of the therapist.

For patients suffering from TRS, an immersive environment such as the virtual reality environment provided by AT helps the therapists in understanding their relationship with their auditory hallucinations in an in vivo setting [41]. This provides insight as to how they may react under specific conditions. The dyadic interaction (Avatar: Coping mechanisms, Patient: Prevention) enables the Avatar to directly elicit prevention strategies from the patient to better understand their cognitive processes in a particular context. Self-generated coping strategies are important in the treatment of psychotic symptoms and these strategies can be assessed through this particular dyad, which is why it might be as prevalent in AT [42].

It is interesting to note that only five dyadic interactions were identified as having a mean frequency above 10 across the participants. This can be hypothesized to be linked by the interpersonal differences across the patients and their own personal experiences with their auditory hallucinations. The therapist personal traits and disposition can affect the therapeutic alliance whereas the patient's own personality traits can also affect it [43]. With the current shift towards precision medicine, it is not surprising to see personalized adaptation of psychotherapy for specific patients [44]. AT is no exception to this, as patients presenting with TRS can be widely heterogenous, and therefore, the interpersonal dynamics can yield multiple different dyadic interactions.

Limitations

This remains an exploratory content analysis and cannot be generalized outside the scope of AT. Considering that two therapists were involved in conducting AT for these patients, the interactions cannot be interpreted as generalizable over all therapists, although there was a consistency observed between all participants and the dyads that emerged. While the amount of verbatims that was analyzed is substantial, the automated annotation has a recurrent bias in its classification, as themes that performed poorly during training due to a lack of data in the initial AT dataset will be misidentified. However, it can be hypothesized that themes that do not occur frequently in AT immersive sessions would signify a smaller mean frequency of apparitions for dyadic interactions that would encompass those themes.

Conclusions

In conclusion, the main goal of this study was to identify the dyadic interactions between the patient and the Avatar that are the most prevalent in AT for patient's suffering from treatment-resistant schizophrenia. The automated annotation of the verbatims and the content analysis of the identified dyadic interactions highlighted five major dyads in AT: (Avatar: Reinforcement, Patient: Self-affirmation), (Avatar: Provocation, Patient: Self-affirmation), (Avatar: Coping mechanisms, Patient: Prevention), (Patient: Self-affirmation, Avatar: Reinforcement), and (Patient: Self-appraisal, Avatar: Reinforcement). These dyads provide a first insight as to the interpersonal dynamics occurring in AT. Future studies on the implication of such dyadic interactions with the therapeutic outcome of AT should be conducted considering the importance of dyadic relationships in psychotherapy.

Author Contributions

Conceptualization, A.H., J.C., M.B., K.P., S.P. and A.D.; methodology, A.H., J.C., L.D., M.B. and A.D.; validation, A.H., J.C. and A.D.; formal analysis, A.H.; investigation, A.H.; data curation, A.H. and J.C.; writing—original draft preparation, A.H.; writing—review and editing, A.H., L.D., M.B., S.P. and A.D.; supervision, K.P., S.P. and A.H.; project administration, K.P.; funding acquisition, K.P., S.P. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work was indirectly supported by Le Fonds de recherche du Québec—Santé (FRQS); Otsuka Canada Pharmaceutical Inc.; Chaire Eli Lilly Canada de recherche en schizophrénie; Fondation Pinel; Services et recherches psychiatriques AD; Fonds d'excellence en recherche Apogée Canada (Institut de la Valorisation des Données IVADO).

Institutional Review Board Statement

This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients. Patients that are part of this study were selected based on the proof-of-concept trial from Percy du Sert 2018’s study and Dellazizzo 2021’s study [28,29]. The trial was conducted in accordance with the Declaration of Helsinki and was approved by the institutional ethical committee (CER IPPM 16-17-06). We obtained written informed consent from all patients.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to patients’ privacy.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Weissman M, Cuijpers P. Psychotherapy over the Last Four Decades. Harv Rev Psychiatry. 2017;25(4):155-8.

2. Locher C, Meier S, Gaab J. Psychotherapy: A World of Meanings. Front Psychol. 2019;10:460.

3. Brent DA, Kolko DJ. Psychotherapy: definitions, mechanisms of action, and relationship to etiological models. J Abnorm Child Psychol. 1998;26(1):17-25.

4. Strauman TJ, Goetz EL, Detloff AM, MacDuffie KE, Zaunmuller L, Lutz W. Self-regulation and mechanisms of action in psychotherapy: a theory-based translational perspective. J Pers. 2013;81(6):542-53.

5. Tzur Bitan D, Shalev S, Abayed S. Therapists' Views of Mechanisms of Change in

Psychotherapy: A Mixed-Method Approach. Front Psychol. 2022;13:565800.

6. Bachelor A. Clients' and therapists' views of the therapeutic alliance: similarities, differences and relationship to therapy outcome. Clin Psychol Psychother. 2013;20(2):118-35.

7. Parth K, Datz F, Seidman C, Loffler-Stastka H. Transference and countertransference: A review. Bull Menninger Clin. 2017;81(2):167-211.

8. Zinn WM. Transference phenomena in medical practice: being whom the patient needs. Ann Intern Med. 1990;113(4):293-8.

9. Jones AC. Transference and countertransference. Perspect Psychiatr Care. 2004;40(1):13-9.

10. Prasko J, Ociskova M, Vanek J, Burkauskas J, Slepecky M, Bite I, et al. Managing Transference and Countertransference in Cognitive Behavioral Supervision: Theoretical Framework and Clinical Application. Psychol Res Behav Manag. 2022;15:2129-55.

11. Hayes JA, Gelso CJ, Goldberg S, Kivlighan DM. Countertransference management and effective psychotherapy: Meta-analytic findings. Psychotherapy (Chic). 2018;55(4):496-507.

12. Prasko J, Diveky T, Grambal A, Kamaradova D, Mozny P, Sigmundova Z, et al. Transference and countertransference in cognitive behavioral therapy. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2010;154(3):189-97.

13. Polese D, Fornaro M, Palermo M, De Luca V, de Bartolomeis A. Treatment-Resistant to Antipsychotics: A Resistance to Everything? Psychotherapy in Treatment-Resistant Schizophrenia and Nonaffective Psychosis: A 25-Year Systematic Review and Exploratory Meta-Analysis. Front Psychiatry. 2019;10:210.

14. Tandon R, Gaebel W, Barch DM, Bustillo J, Gur RE, Heckers S, et al. Definition and description of schizophrenia in the DSM-5. Schizophr Res. 2013;150(1):3-10.

15. McCutcheon RA, Abi-Dargham A, Howes OD. Schizophrenia, Dopamine and the Striatum: From Biology to Symp-toms. Trends Neurosci. 2019;42(3):205-20.

16. Weinstein JJ, Chohan MO, Slifstein M, Kegeles LS, Moore H, Abi-Dargham A. Pathway-Specific Dopamine Abnor-malities in Schizophrenia. Biol Psychiatry. 2017;81(1):31-42.

17. Correll CU, Schooler NR. Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, As-sessment, and Treatment. Neuropsychiatr Dis Treat. 2020;16:519-34.

18. Elkis H, Buckley PF. Treatment-Resistant Schizophrenia. Psychiatr Clin North Am. 2016;39(2):239-65.

19. Miyamoto S, Jarskog LF, Fleischhacker WW. New therapeutic approaches for treatment-resistant schizophrenia: a look to the future. J Psychiatr Res. 2014;58:1-6.

20. Nucifora FC, Jr., Woznica E, Lee BJ, Cascella N, Sawa A. Treatment resistant schizophrenia: Clinical, biological, and therapeutic perspectives. Neurobiol Dis. 2019;131:104257.

21. Lally J, Gaughran F. Treatment resistant schizophrenia - review and a call to action. Ir J Psychol Med. 2019;36(4):279-91.

22. Rakitzi S, Georgila P. Integrated Psychological Therapy and Treatment-Resistant Schizophrenia: Initial Findings. Psychiatry. 2019;82(4):354-67.

23. Morrison AP, Pyle M, Gumley A, Schwannauer M, Turkington D, MacLennan G, et al. Cognitive-behavioural therapy for clozapine-resistant schizophrenia: the FOCUS RCT. Health Technol Assess. 2019;23(7):1-144.

24. Morrison AP, Pyle M, Gumley A, Schwannauer M, Turkington D, MacLennan G, et al. Cognitive behavioural therapy in clozapine-resistant schizophrenia (FOCUS): an assessor-blinded, randomised controlled trial. Lancet Psychiatry. 2018;5(8):633-43.

25. Burns AM, Erickson DH, Brenner CA. Cognitive-behavioral therapy for medication-resistant psychosis: a me-ta-analytic review. Psychiatr Serv. 2014;65(7):874-80.

26. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff AP. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? Psychosis. 2014;6(2):166-76.

27. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hal-lucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry. 2018;5(1):31-40.

28. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophr Res. 2018;197:176-81.

29. Dellazizzo L, Potvin S, Phraxayavong K, Dumais A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. NPJ Schizophr. 2021;7(1):9.

30. Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychother. 2018;25(6):878-85.

31. Beaudoin M, Potvin S, Machalani A, Dellazizzo L, Bourguignon L, Phraxayavong K, et al. The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clin Psychol Psychother. 2021;28(3):500-18.

32. Morina N, Brinkman WP, Hartanto D, Kampmann IL, Emmelkamp PM. Social interactions in virtual reality exposure therapy: A proof-of-concept pilot study. Technol Health Care. 2015;23(5):581-9.

33. Hudon A, Beaudoin M, Phraxayavong K, Dellazizzo L, Potvin S, Dumais A. Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach. Health Informatics J. 2022;28(4):14604582221142442.

34. Butler EA. Interpersonal Affect Dynamics: It Takes Two (and Time) to Tango. Emotion Review. 2015;7(4):336-41.

35. Petrocchi S, Iannello P, Lecciso F, Levante A, Antonietti A, Schulz PJ. Interpersonal trust in doctor-patient relation: Evidence from dyadic analysis and association with quality of dyadic communication. Soc Sci Med. 2019;235:112391.

36. Markin RD, McCarthy KS, Barber JP. Transference, countertransference, emotional expression, and session quality over the course of supportive expressive therapy: the raters' perspective. Psychother Res. 2013;23(2):152-68.

37. Blair RJR. Traits of empathy and anger: implications for psychopathy and other disorders associated with aggression. Philos Trans R Soc Lond B Biol Sci. 2018;373(1744).

38. Corrigan PW, Green MF. Schizophrenic patients' sensitivity to social cues: the role of abstraction. Am J Psychiatry. 1993;150(4):589-94.

39. Nikolaides A, Miess S, Auvera I, Muller R, Klosterkotter J, Ruhrmann S. Restricted attention to social cues in schiz-ophrenia patients. Eur Arch Psychiatry Clin Neurosci. 2016;266(7):649-61.

40. Chaix J, Ma E, Nguyen A, Ortiz Collado MA, Rexhaj S, Favrod J. Safety-seeking behaviours and verbal auditory hal-lucinations in schizophrenia. Psychiatry Res. 2014;220(1-2):158-62.

41. Bisso E, Signorelli MS, Milazzo M, Maglia M, Polosa R, Aguglia E, et al. Immersive Virtual Reality Applications in Schizophrenia Spectrum Therapy: A Systematic Review. Int J Environ Res Public Health. 2020;17(17).

42. Ruckl S, Gentner NC, Buche L, Backenstrass M, Barthel A, Vedder H, et al. Coping with delusions in schizophrenia and affective disorder with psychotic symptoms: the relationship between coping strategies and dimensions of delusion. Psy-chopathology. 2015;48(1):11-7.

43. Chapman BP, Talbot N, Tatman AW, Brition PC. Personality Traits and the Working Alliance in Psychotherapy Trainees: An Organizing Role for the Five Factor Model? J Soc Clin Psychol. 2009;28(5).

44. Barber JP, Solomonov N. Toward a personalized approach to psychotherapy outcome and the study of therapeutic change. World Psychiatry. 2019;18(3):291-2.

Figures and Tables

Table 1. Avatar interactions themes as per Beaudoin et al. 2021.

Avatar themes	**Examples**
Accusations	“You are responsible for this. “
Omnipotence	“I am the best.’’
Beliefs	“I believe that you are ill.’’
Active listening, empathy	“There is no rush, take your time. ’’
Incitements, orders	“You should hit yourself.’’
Coping mechanisms	“Why are you not happy when I insult you? ’’
Threats	“I will kill you’’.
Negative emotions	“It’s difficult for me to realize that.’’
Self-perceptions	“I see myself as worthless.’’
Positive emotions	“I am feeling great.’’
Provocation	“Try me. ’’
Reconciliation	“Should we stop arguing?’’
Reinforcement	“You should do this again. ’’

Table 2. Patient interactions’ themes as per Beaudoin et al. 2021.

Patient themes	**Examples**
Approbation	“You are right’’
Self-deprecation	“I can’t do this.’’
Self-appraisal	“I am a nice person.’’
Other beliefs	“You are the one controlling me’’
Counterattack	“You are the one who did this, not me!’’
Maliciousness of the voice	“You are trying to make this hard for everyone.’’
Negative	“It is not easy.’’
Negation	“I did not do this.’’
Omnipotence	“I am everywhere.’’
Disappearance of the voice	“Please vanish!’’
Positive	“I am feeling great.’’
Prevention	“I will try not pay attention to you.’’
Reconciliation of the voice	“Let`s be friends.’’
Self-affirmation	“I can do this.’’

Table 3. Classification performances for the automated annotation of the AT verbatims.

Avatar Themes	**Precision**	**Recall**	**f1-score**	**#Interactions tested**
Accusations	0.74	0.66	0.70	35
Omnipotence	0.77	0.94	0.85	18
Beliefs	0.75	0.69	0.72	26
Active listening, empathy	0.70	0.82	0.76	17
Incitements, orders	0.78	0.70	0.74	10
Coping mechanisms	0.96	0.88	0.92	25
Threats	1.00	0.86	0.92	7
Negative emotions	0.79	0.92	0.85	12
Self-perceptions	0.60	0.71	0.65	17
Positive emotions	0.91	0.62	0.74	16
Provocation	0.67	0.62	0.65	16
Reconciliation	0.64	0.69	0.67	13
Reinforcement	0.73	0.84	0.78	19
Accuracy			0.76	231
Weighted average	0.77	0.76	0.76	231
Patient Themes	**Precision**	**Recall**	**f1-score**	**#Interactions tested**
Approbation	0.31	0.31	0.31	13
Self-deprecation	0.53	0.67	0.59	12
Self-appraisal	0.86	0.59	0.70	32
Other beliefs	0.52	0.65	0.58	17
Counterattack	0.70	0.54	0.61	26
Maliciousness of the voice	0.62	0.71	0.67	14
Negative	0.72	0.64	0.68	36
Negation	0.79	0.73	0.76	30
Omnipotence	0.27	0.60	0.37	10
Disappearance of the voice	0.78	0.64	0.70	22
Positive	0.83	0.88	0.86	17
Prevention	0.86	0.59	0.70	32
Reconciliation of the voice	0.30	1.00	0.46	3
Self-affirmation	0.46	0.62	0.53	21
Accuracy			0.64	285
Weighted average	0.69	0.64	0.65	285

Table 4. Patients' characteristics

Characteristics	**Value (N=32)**
Sex (male, female)	24,8
Age (mean in years)	42.6 ± 11.0
Education (mean in years)	13.6 ± 3.0
Ethnicity (Caucasian, others)	93.4%,6.6%
% on Clozapine	40.0%

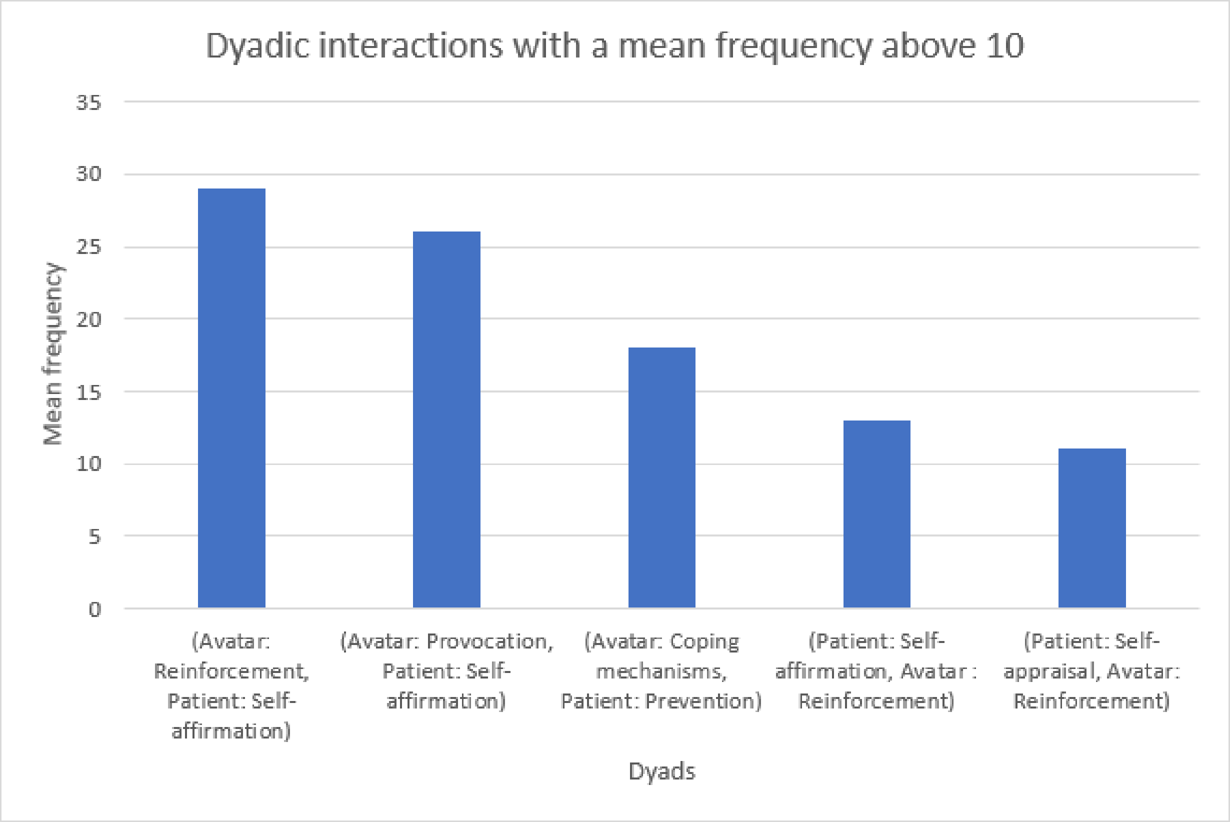

Figure 1. Dyadic interactions and their mean frequencies. Only dyads with a mean frequency per participants above 10 over the entire AT were presented.

Chapitre 4 – DISCUSSION

Synthèse des principaux résultats

L'objectif principal de cette thèse était d'aider les thérapeutes dans l'amélioration de la réponse thérapeutique de patients souffrant de schizophrénie avec des hallucinations auditives réfractaires à la médication, en intégrant les principes de l'intelligence artificielle dans la TA pour la prédiction de la réponse clinique des patients. Ceci est réalisé en utilisant comme pistes de réflexion les résultats des dix études présentées autour de cinq sous-objectifs : (i) utiliser l'intelligence artificielle pour prédire les issues cliniques des patients atteints de troubles mentaux graves et persistants (étude 1) ; (ii) évaluer les possibilités d'intégration de l'intelligence artificielle dans la psychothérapie sur les plans technique et clinique (études 2 et 3) ; (iii) utiliser l'apprentissage automatique dans la TA pour l'annotation automatisée des séances thérapeutiques immersives (études 4 à 7) ; (iv) intégrer les algorithmes d'apprentissage automatique pour prédire la réponse clinique des patients suivant la TA (étude 8) ; (v) identifier les facteurs prédictifs multimodaux qui pourraient améliorer la prédiction de l'issue clinique des patients (études 9 et 10).

La première étude a démontré qu'il était possible d'utiliser un modèle de prédiction pour anticiper l’issue clinique de la violence chez des patients atteints de troubles mentaux graves et persistants consommateurs de cannabis. Les revues de littérature (études 2 et 3) ont permis d'évaluer les possibilités d'intégration de l'intelligence artificielle dans la psychothérapie, en identifiant cinq algorithmes d'apprentissage automatique utilisés pour de petites bases de données et en illustrant trois grands champs d'application clinique : la prédiction des résultats thérapeutiques des patients, l'analyse de contenu de la psychothérapie et la catégorisation automatique des interactions psychothérapeutiques. Les résultats des études 4 à 7 ont mis en relief plusieurs algorithmes précédemment identifiés comme ayant du potentiel pour la TA. L'algorithme LSVC a été évalué comme étant le plus performant parmi les approches supervisées, avec une performance comparable à celle d'une approche par méthode d'ensemble pour la TA et la base de données actuelle sur la TA. Une approche non supervisée de type KNN a également démontré une linéarité dans les interactions de la TA, comparable aux études précédemment réalisées par Dellazizzo et al. (2018) et Beaudoin et al. (2021). L'étude 8 a permis d'intégrer un algorithme de type LSVC pour annoter automatiquement des interactions issues de la TA et de prédire l'évolution des patients en ce qui concerne leurs hallucinations auditives, en combinant cet algorithme à un algorithme de régression logistique. Finalement, les résultats des études 9 et 10 ont permis de mieux comprendre les émotions et les dyades thérapeutiques intervenant dans la TA, qui pourraient être intégrées comme des facteurs multimodaux dans la prédiction de la réponse des patients à cette thérapie.

i. Utiliser l'intelligence artificielle dans la prédiction des issues cliniques des patients atteints de troubles mentaux graves et persistants

Utilisation de l’apprentissage machine dans la prédiction de la violence

La première étude a démontré qu’il était possible de modéliser, à l’aide d’un algorithme de type régression logistique LASSO, le risque de violence associé à divers facteurs dont la consommation de cannabis chez des patients atteints de troubles mentaux graves et persistants. Ces facteurs incluent par exemple : l'utilisation du cannabis au cours des 12 derniers mois après avoir décidé qu'il serait préférable de s'abstenir, avoir des difficultés à passer une journée sans consommer, et considérer que le temps passé à utiliser le cannabis est problématique ; ces éléments sont des prédicteurs positifs forts liés au cannabis pour les comportements violents (Hudon et al., 2023). Les modèles de régression dans la prédiction d’une issue clinique sont de plus en plus utilisés et émergents dans la recherche en santé mentale, avec peu d’application actuellement en clinique (Meehan et al., 2022). Cette problématique est probablement liée au fait que, malgré l’utilisation et le perfectionnement des modèles mathématiques, notamment ceux issus de l’apprentissage machine, de nombreux défis et limites continuent d’être inhérents. Ces enjeux limitent donc la validité externe de ces modèles lorsqu’ils sont utilisés dans d’autres contextes. Par exemple, dans le contexte du trouble dépressif caractérisé, l’équipe de Ermers et al. (2020) a réalisé une revue de la littérature afin d’examiner l'utilisation de modèles d'apprentissage automatique pour prédire les résultats du traitement (pharmacologique) et d'identifier des sous-groupes de patients au sein du trouble dépressif majeur. Les auteurs soulignent que les modèles prédictifs, qui sont principalement au stade de preuve de concept, démontrent une précision prédictive modérée avec des valeurs d'aire sous la courbe allant de 0,63 à 0,78 (substantiel à bon) (Ermers et al., 2020). Ils mentionnent que les principales limites de ces modèles sont le nombre de données utilisées et le contexte clinique, ainsi qu'une absence de validité externe (tests avec des échantillons qui ne sont pas tirés de leurs données pour entraîner les modèles). L'enjeu des données est rapporté dans notre première étude, puisqu'il est difficile d'avoir une base de données importante pour les troubles mentaux graves et persistants, ce qui a limité les performances de l'algorithme, d'autant plus qu'il n'y avait pas de base de données comparatives, considérant la nature transversale de l'étude (Hudon et al., 2023).

Toutefois, une des forces de cette première étude est que les prédicteurs mis en évidence par l'algorithme se comparent à d'autres études dans le domaine qui ont utilisé d'autres modalités méthodologiques afin d'explorer les prédicteurs de violence chez cette population. Ceci est encourageant, puisque la revue de littérature la plus récente sur l'application de l'apprentissage machine pour la schizophrénie et le trouble bipolaire rapporte (à partir d'une analyse de 15 études sur un échantillon de 1243 articles) que, même s'il est difficile de comparer les modèles entre eux, l'apprentissage automatique peut prédire avec précision les résultats et aider à prendre des décisions cliniques concernant la schizophrénie et le trouble bipolaire lorsqu'utilisés dans un contexte précis avec des données pertinentes (Montazeri et al., 2022). L'utilisation de facteurs de risques d'impulsivité, de consommation et de comorbidités psychiatriques est donc un choix judicieux lors de la conception de modèle prédictif portant sur le risque de violence, car cela s'accorde a priori avec la littérature sur le sujet.

Un autre enjeu important à considérer, et qui a été soulevé dans cette première étude, est qu'il est très difficile de prédire les comportements humains. Dans leur perspective portant sur l'application de modèles pour prédire le comportement humain, Kleinberg et al. (2023) soulignent un problème courant où les algorithmes prédisent le comportement dans l'intention de déduire des états mentaux. Les auteurs soutiennent que cette approche peut être problématique en raison de l'écart entre les comportements observés et les états mentaux réels des patients pour lesquels on tente de prédire les comportements (Kleinberg et al., 2023). Ceux-ci soulignent également la nécessité pour les algorithmes de considérer les perspectives psychologiques afin de corréler avec précision les états mentaux à partir des comportements observés, au lieu de se fier uniquement aux prédictions comportementales (Kleinberg et al., 2023). Cela apporte donc une autre vision quant à la performance prédictive de l'algorithme LASSO, avec un $R^2 = 0,41$, qui est conforme aux scores prédictifs trouvés dans la littérature pour la prédiction de différents comportements humains (Hamilton et al., 2015).

Conclusion de la section

Cette première étude s'accorde avec la littérature scientifique dans le fait qu'il est possible d'utiliser des modèles de prédictions d'issues cliniques pour les personnes atteintes d'un trouble mental grave et persistant (dans ce contexte pour prédire la violence chez les utilisateurs de cannabis issus de cette population). Cependant, ces modèles ne sont pas sans failles et s'accompagnent de plusieurs considérations quant à l'interprétation de leurs résultats. Le constat concernant les limites intrinsèques à la performance des algorithmes pour la prédiction des comportements pourrait être amélioré si on prenait davantage en compte les états mentaux des patients. Ceci sous-entend donc que l'utilisation de variables qui reflètent davantage les états mentaux, telles que les interactions thérapeutiques dans le cadre d’une psychothérapie, pourrait s'avérer bénéfique pour augmenter la puissance prédictive des modèles mathématiques utilisés dans un tel contexte.

ii. Évaluer les possibilités d'intégration de l'intelligence artificielle dans la psychothérapie sur les plans techniques et cliniques

Recension des écrits sur l’utilisation de l’apprentissage machine sur de petites bases de données dans un contexte pour l’annotation automatisée de thérapies

Cette deuxième étude a permis d’identifier les algorithmes d’apprentissage machine utilisés dans le contexte de petites bases de données, telles que dans les domaines de la psychiatrie, de la psychothérapie et des sciences sociales, pour la classification automatisée d’entités textuelles. Sans surprise, très peu d’articles ont été identifiés dans le cadre de cette revue de la littérature, puisque les assises actuelles quant à l’utilisation de modèles prédictifs reposent sur l’utilisation de bases de données habituellement très larges (Riley et al., 2016). Cette revue systématique a révélé que les classificateurs de type support machine vectoriel sont les plus précis pour la classification automatique de textes dans de petits ensembles de données dans les domaines à l’étude, notamment lorsque les données sont linéaires (Hudon et al., 2021). Dans le cadre de cette étude, sept articles ont été analysés, identifiant que les scores de classification prédictive des algorithmes utilisés dans ces articles variaient entre 53 % et 91 % pour classer des entités textuelles en 4 à 7 catégories (Hudon et al., 2021). Le degré de validation des algorithmes n’était pas comparable, considérant l’hétérogénéité des issues étudiées dans ces articles. Par ailleurs, seulement trois des études examinées ont rapporté des statistiques d'accord inter-juges, qui se sont avérées cohérentes avec les statistiques d'accord inter-juges pour les classifications de texte réalisées par des évaluateurs humains (Hudon et al., 2021).

Les algorithmes de type support machine vectoriel ont tendance à mieux performer dans la classification d’entités textuelles pour de petites bases de données et pour les échantillons plus larges (Hu et al., 2022). À cet effet, la plupart des études actuelles portant sur les patients atteints de schizophrénie, pour ce qui est de la classification automatisée de matériel à l’aide de l’intelligence artificielle, portent sur des données biologiques, des imageries et des résultats d’électroencéphalogrammes (Verma et al., 2023). Toutefois, pour la classification d’entités textuelles, cette deuxième étude devient d’actualité, notamment dans le contexte de la psychothérapie, puisque les études axées sur mieux comprendre le contenu thérapeutique sont souvent réalisées à l’aide de modalités qualitatives afin de faire émerger les phénomènes thérapeutiques (Palinkas, 2015). Comme recensé dans cette étude systématique, très peu de chercheurs se sont aventurés dans l’analyse d’entités textuelles en apprentissage machine, en raison des enjeux de données qui sont souvent limités et disparates. Une revue de la littérature sur le sujet a permis de mettre en évidence des pistes de solutions pour améliorer les données issues de la psychothérapie afin de favoriser l’utilisation de l’apprentissage machine pour classifier le texte analysé. Dans cette étude de Smink et al. (2019), les principaux résultats ont indiqué que le domaine de la recherche sur le processus de changement thérapeutique est fragmenté, avec quelques méthodes fréquemment utilisées montrant un potentiel d'automatisation dans l'analyse des séances de thérapie. Ces méthodes, particulièrement prometteuses pour l'automatisation, comprenaient le *Schéma de Codage des Moments Innovants*, l'*Échelle d'Assimilation des Expériences Problématiques*, le *Schéma de Codage du Processus Narratif* et l'*Analyse de Conversation*, se concentrant principalement sur l'amélioration du patient et les interactions thérapeute-patient (Smink et al., 2019). Dans ce contexte, raffiner le processus de classification en se concentrant sur une modalité, telle que l’analyse conversationnelle, pourrait améliorer la qualité des données utilisées dans le modèle de classification et permettre une meilleure performance de ce dernier. Cela expliquerait probablement, si une analyse qualitative sur laquelle se base le modèle prédictif a été réalisée afin de s’assurer de la linéarité des interactions à classifier (chaque interaction se classe facilement dans une seule catégorie), qu’un modèle comme le LSVC performe mieux vu son optimisation pour les données qui ont une valence linéaire.

Recension des écrits sur l’intégration des réseaux neuronaux dans la psychothérapie

Cette troisième étude a permis d’identifier les utilisations d’algorithmes de type réseaux neuronaux dans le contexte de la psychothérapie clinique. Les études appliquant des algorithmes de réseaux neuronaux dans un contexte psychothérapeutique ont été incluses dans la revue de littérature. Parmi les principaux résultats, trois applications principales des réseaux neuronaux ont été identifiées : prédire les résultats psychothérapeutiques, analyser le contenu des thérapies et automatiser la catégorisation des interactions psychothérapeutiques (Hudon et al., 2023).

Cette étude met également en évidence le peu d'études sur le sujet. En revanche, le domaine de la psychiatrie clinique a davantage exploré ce concept de prédiction d’issues cliniques à l’aide de réseaux neuronaux dans des contextes très spécifiques. Un exemple est l’évaluation de l’anxiété et de l’espoir des patients atteints de cancer à l’aide de marqueurs biométriques tels que l’analyse de séquences visuelles (Shafiei et al., 2020). Les auteurs d’une revue d’experts sur l’utilisation des réseaux neuronaux en psychiatrie rapportent que ces algorithmes sont particulièrement utiles dans d'autres domaines de la médecine qui dépendent de l'analyse d'images, tels que la détection de tumeurs en radiologie, et montrent quelques promesses dans le diagnostic psychiatrique, aidant à former des décisions diagnostiques plus standardisées à l’aide de données quantitatives ciblées (Durstewitz et al., 2019). Toutefois, la majorité des études en psychiatrie utilisant les réseaux neuronaux qu’ils ont identifiées se sont concentrées sur l’aide au diagnostic et la prédiction des trajectoires de maladie. L'intégration de données multimodales, combinant des informations d'imagerie neurologique et de génomique, par exemple, a montré que les réseaux neuronaux surpassent les méthodes traditionnelles dans les décisions diagnostiques (Durstewitz et al., 2019). Cela explique probablement pourquoi très peu d’études ont été recensées dans le domaine de la psychothérapie, considérant le peu d’analyses multimodales effectuées dans ce contexte et les données qualitatives qui ne sont habituellement pas celles utilisées par ces algorithmes qui nécessitent de grands échantillons de données (Emmert-Streib et al., 2020).

Considérant les analyses de contenu des thérapies et l’automatisation des interactions psychothérapeutiques par les réseaux neuronaux dans un contexte de psychothérapie, les études manquent également à l’appel. Les efforts de recherche sont davantage axés sur l’analyse textuelle de grands échantillons de données (par exemple : dossiers médicaux, médias sociaux, entrevues) d’amalgames de patients, plutôt que sur des transcrits de psychothérapies, puisque l’analyse du langage naturel nécessite habituellement de vastes échantillons de données dans un contexte clinique spécifique (Zhang et al., 2022). De plus, dans la classification automatisée d’entités textuelles, le concept de sémantique est important et les algorithmes disponibles en accès libre ont souvent des lexiques optimisés principalement pour l’anglais, ce qui limite l’application à d’autres contextes culturels (Sarica et al., 2021), expliquant ainsi pourquoi on voit peu d’applications des réseaux neuronaux dans l’analyse de verbatims pour la psychothérapie.

Conclusion de la section

Ces deux études ont permis de délimiter les algorithmes d'intelligence artificielle qui pourraient être intégrés dans la psychothérapie sur les plans techniques, notamment pour la classification automatisée d'interactions thérapeutiques à l'aide d'entités textuelles. La deuxième étude a permis de circonscrire les algorithmes qui peuvent avoir des performances satisfaisantes dans le contexte de petites bases de données. Elle a donc bien mis en évidence que les arbres décisionnels, les classificateurs de type support machine vectoriel et les algorithmes naïfs bayésiens sont des algorithmes potentiellement utilisables pour l'intégration dans la classification d'interactions thérapeutiques pour une thérapie telle que la TA. Le troisième article a permis d'identifier les applications d'une famille d'algorithmes plus performants, mais qui demandent souvent un degré d'optimisation important ainsi qu'un grand nombre de données, rendant ces algorithmes difficilement utilisables dans le cadre de la TA.

iii. Utiliser l'apprentissage machine dans la thérapie par Avatar pour l'annotation automatisée des séances thérapeutiques immersives

Comparaison de différents algorithmes pour annoter automatiquement les séances immersives de la thérapie Avatar

La quatrième étude s'est penchée sur la comparaison des performances de classification automatique, des interactions thérapeutiques prenant place dans la TA, de cinq algorithmes d'apprentissage automatique : le SVM, le LSVC, le classificateur naïf bayésien multinomial, le classificateur d'arbre décisionnel, et le classificateur de perceptron multicouche. Ces derniers ont été testés sur un ensemble de données issues de transcriptions de séances immersives de TA de 35 patients pour automatiser l'annotation des interactions thérapeutiques. L'étude a révélé que le classificateur linéaire à vecteurs de support surpassait les autres algorithmes en termes de précision, de rappel et de score F1, tandis que le classificateur à vecteurs de support régulier était le meilleur pour la précision (Hudon et al., 2023).

Dans cette étude, les algorithmes qui ont été comparés était ceux qui ont été identifié dans la deuxième étude. Déjà à ce stade, les algorithmes de type SVM (tel que le classificateur SVM et le LSVC) avaient été identifiés comme étant plus performants pour les entités textuelles issues de petites bases de données (Hudon et al., 2021). La performance du LSVC, dans le cadre de son application à la TA, peut être expliqué par la linéarité des interactions thérapeutiques qui ont déjà été pré-étudiées au préalable dans le cadre des études qualitatives antérieures portant sur la TA (Dellazizzo et al., 2018 ; Beaudoin et al, 2021). Dans leur étude sur des bases de données textuelles, l'équipe de Krakovska et al. (2019) que les algorithmes de types LSVC ont tendance à mieux performer lorsque les données sont linéaires avec le moins de codépendance possible. Lorsque ce n'est pas le cas, les algorithmes qui utilisent des noyaux d'exploitation (kernel) non-linéaires ont tendance à mieux performer (Krakovska et al., 2019).

Toutefois, il est important de prendre en compte que dans cette étude comparative, plusieurs limites étaient inhérentes à la petite base de données utilisées et le contexte très spécifique de la TA. Des algorithmes comme le classificateur de perceptron multicouche, nécessite beaucoup de données et donc la performance de ces algorithmes pourraient s'améliorer avec le temps puisque de nouvelles données seront ajoutées à cette base de données (Percha et al., 2021). Il faut donc prendre en considération que le choix d'algorithme est dynamique et doit être contextualisé afin d'optimiser les performances de celui-ci, mais également de la qualité des données utilisées (par exemple : linéaires versus non-linéaires).

Implémentation d'un classificateur de type support vectoriel linéaire pour annoter les séances immersives de la thérapie par Avatar

La cinquième étude a permis d'optimiser l'algorithme LSVC identifié dans la quatrième étude pour classer automatiquement les interactions thérapeutiques retrouvées dans les transcriptions des séances immersives de la TA. Un ensemble de données de 162 transcriptions, annotées manuellement avec 28 thèmes, pour entraîner et tester l'algorithme a permis d'atteindre une précision de classification globale de 66,02% et un accord de classification substantiel de 0,647 par rapport aux codeurs humains (Hudon et al., 2022). Ces résultats suggèrent que l'apprentissage automatique peut efficacement aider à l'analyse qualitative des séances de psychothérapie en offrant un moyen plus efficace d'analyser les processus thérapeutiques.

Tel que mentionné précédemment, la nature linéaire des données de la TA ont permis d’obtenir une performance comparable aux classifications de codeurs humains lorsque comparé aux travaux de Dellazizzo et al. (2018) et Beaudoin et al. (2021). L’étude soulève également la comparaison avec de nombreuses études issues de la littérature qui démontrent des performances comparables pour la classification automatisée de texte dans un contexte médical pour de petite bases de données. La question demeure quant à la possibilité d’obtenir une performance prédictive plus grande que celle obtenue dans cette étude et sa validité externe.

En lien avec la performance prédictive, plusieurs stratégies sont identifiées dans la littérature. Par exemple, il y a plus d’une décade, Zhang et al. (2008) ont mis en évidence des stratégies afin d’optimiser les performances de classificateurs de type LSVC pour la classification automatisée du texte telles que : la stratégie de décomposition (modifier une longue séquence de mots pour plusieurs petites séquences sur un même thème), la stratégie de combinaison (contextualiser une séquence de mots dans une phrase ou une entité sémantique) et l’élimination de mots à faible valence dans le contexte sémantique (par exemple : enlever les déterminants dans les phrases) (Zhang et al., 2008). Ces stratégies n’ont pas pu être mises en place dans cette cinquième étude puisque la réalité de la base de données issue des séances immersives de la TA se déroule dans un contexte francophone, québécois, avec une grande hétérogénéité quant à l’expérience des patients en lien avec leur voix la plus menaçante ce qui permet difficilement d’en décomposer ou combiner le contexte. Quant à l’élimination de mots à faible valence, au moment de la réalisation de cette étude, il n’y avait pas de dictionnaire québécois disponible pour utilisation comme paramètre de l’algorithme LSVC, rendant cette tâche ardue. Avec un dictionnaire, ou du moins une séquence de mots validées, en français québécois, il pourrait être possible d’optimiser la vectorisation des interactions thérapeutiques et potentiellement obtenir une meilleure performance prédictive.

Utilisation de l’apprentissage machine non-supervisé pour annoter automatiquement les interactions de la thérapie par Avatar

Cette sixième étude avait pour but d’utiliser l'apprentissage automatique non supervisé (en particulier, un algorithme de regroupement K-means) pour analyser les interactions thérapeutiques des séances de thérapie de 18 patients suivant la TA Cette analyse visait principalement à catégoriser les interactions en différents clusters et à les comparer avec des thèmes identifiés précédemment à partir des analyses qualitatives de Dellazizzo et al. (2018) et Beaudoin et al. (2021). Les principaux résultats ont révélé trois clusters pour les interactions de l'avatar et quatre pour les patients, démontrant le potentiel de l'apprentissage automatique pour fournir des aperçus quantitatifs sur la dynamique des séances de TA.

Les résultats de cette étude sont importants puisque les clusters identifiés par l’algorithme d’apprentissage machine non-supervisé à permis de mettre en évidence une corrélation avec les études précédentes quant à la classification des interactions thérapeutiques. Cela ajoute donc à l’argument de la linéarité des données présents dans la banque de données de la TA. Ce type d’approche permet également d’identifier des liens entre des groupes de données et d’évaluer la force de ces liens. Lorsque la base de données de la TA sera plus importante, il est possible de voir une certaine migration des données puisque de nouveaux types d’interactions pourraient être identifiés (considérant la complexité des patients présentant une SRT et l’hétérogénéité de l’expérience humaine quant au vécu expérientiel face à l’expérience hallucinatoire). À cet effet, une approche non-supervisée pourrait rapporter ces changements de clusters afin de raffiner la classification des interactions thérapeutiques (Nadif et al., 2021).

De plus, ce type d’analyse pourrait permettre l’identification de profil de patients suivant la TA étant donné leurs interactions thérapeutiques avec la TA (Eckhardt et al, 2015). Ceci serait une approche intéressante considérant que la schizophrénie présente une variété importante de présentation clinique (Huang et al., 2020). Un des problèmes réside dans l'interprétation de ces clusters et de comprendre sur quelles assises ils ont été formés avant d'émettre des hypothèses quant à leur validité clinique (Ghassemi et al., 2020).

En lien avec la TA, il sera donc intéressant de continuer de mettre en perspective ces clusters avec l'augmentation des données dans la base de données afin d'observer les changements quant aux regroupements des interactions psychothérapeutiques et l'importance relative des clusters en lien avec la réponse clinique des patients. Considérant l'hétérogénéité des présentations cliniques des patients atteints d'une SRT, l'apprentissage non-supervisé pourrait se révéler être une approche de choix car elle n'est pas susceptible aux biais de classification mis sur pied par l'humain puisqu'elle fait émerger des tendances à partir des données, ce qui est donc une perspective intéressante pour la médecine de précision (Deo, 2015).

Implémentation d'un regroupement de classificateurs pour améliorer la performance de la classification des interactions thérapeutique dans la thérapie par Avatar

Cette septième étude a exploré l'efficacité d'une approche d'ensemble utilisant cinq algorithmes d'apprentissage automatique (SVM, bayes naïf multinomial, classificateur de perceptron multicouche, XGBClassifier et modèle K-Nearest-Neighbors) pour la classification automatique des interactions thérapeutiques dans les verbatims des séances immersives de la TA. Cette méthode a été testée contre des modèles individuels pour améliorer la précision, le rappel, la précision et le score F1 dans l'identification des thèmes d'interaction thérapeutiques des patients et des avatars. Le modèle d'ensemble a surpassé les modèles individuels en termes de précision et de la plupart des métriques, en particulier pour les interactions avec l'avatar, démontrant son potentiel pour une classification de texte plus précise et fiable dans les contextes psychothérapeutiques. Les performances étaient toutefois très similaires au LSVC utilisé de façon individuelle.

Une étude récente portant sur la comparaison d'algorithmes à un ensemble d'algorithme à l'aide de données (libre-accès) portant sur diverses issues en santé mentale à démontrer que de façon générale, les méthodes d'ensemble performent mieux que leurs contreparties et ce autant pour la sensibilité que la spécificité dans le cadre de prédiction du diagnostic clinique (Chung et al., 2023). Dans le contexte de la schizophrénie, Lin et al. (2021) ont utilisé une approche d'apprentissage automatique en ensemble pour prédire des résultats fonctionnels dans la schizophrénie, en utilisant les symptômes cliniques et les fonctions cognitives de 302 patients taïwanais. Ceci impliquait une sélection de caractéristiques à partir de trois échelles de symptômes cliniques et de onze scores de fonctions cognitives, en utilisant l'algorithme de sélection de caractéristiques M5 Prime. Cette approche a été comparée à d'autres modèles d'apprentissage automatique tels que les réseaux neuronaux multicouches, les SVM, la régression linéaire et les forêts aléatoires. Les résultats ont montré que le modèle d'ensemble avec sélection de caractéristiques surpassait les autres modèles dans la prédiction de l'échelle de qualité de vie et de l'évaluation globale du fonctionnement (Lin et al., 2021). De plus, la combinaison de l'échelle d'évaluation des symptômes négatifs (SANS20) et de l'échelle de dépression de Hamilton (HAMD17) était la plus prédictive pour le résultat QLS, tandis que la combinaison de SANS20 et de l'échelle positive et négative du syndrome de PANSS (PANSS-Positive) était la plus prédictive pour le résultat de l'évaluation globale du fonctionne (Lin et al., 2021). Il n'est donc pas surprenant que pour des données issues d'un seul contexte tel que les interactions thérapeutiques dans la TA, un ensemble d'algorithme performe mieux que ses contreparties. Il s'agit également d'une approche de choix lorsqu'on veut favoriser la stabilité des performances dans le temps et leur fiabilité (Naderalvojoud et al., 2024).

Toutefois, considérant la linéarité des données qui sont dans la base de données de la TA, les performances de la méthode d'ensemble se sont révélées quasi-similaires au LSVC. Parmi les enjeux rencontrés avec les méthodes d'ensemble, il y a notamment le choix des modèles à intégrer dans le modèle d'ensemble, le temps d'exécution du modèle et l'interprétation des résultats compte tenu de la complexité de l'ensemble de modèle et des performances des algorithmes qui le composent (Cao et al, 2020). Compte-tenu de la faible supériorité de la performance de l'ensemble comparé au LSVC pour son intégration à la TA, la LSVC demeure un algorithme de choix à ce stade pour la classification automatisée.

Conclusion de la section

Dans le cadre de ces quatre études, il a été mis en évidence que plusieurs algorithmes peuvent s'inscrire dans la classification automatisée d'interaction thérapeutiques prenant place dans les séances immersives de la TA. Les méthodes d'ensemble (regroupement d'algorithmes) peuvent également être utilisées dans ce contexte. Toutefois, c'est le LSVC qui se démarque à ce stade dans le contexte spécifique de la TA avec les données actuelles du fait de ses performances et sa complexité qui est moindre que pour les méthodes d'ensemble. L'analyse non-supervisée des données de la TA a également permis d'identifier les clusters d'interactions des avatars et des patients et d'en apprécier la nature linéaire des interactions. L'utilisation de l'apprentissage machine est donc une modalité qui permet la classification automatisée de contenu thérapeutique dans la TA.

iv. Intégrer les algorithmes d'apprentissage machine pour prédire la réponse clinique des patients suivant la thérapie par Avatar

Combiner un algorithme de classification à un algorithme de régression afin de prédire le devenir clinique des patients suivant la thérapie par Avatar

La huitième étude visait à prédire les résultats thérapeutiques de la TA de patients atteints d'une SRT à partir de leur première séance immersive en utilisant une combinaison de LSVC (pour l'annotation automatisée des interactions thérapeutiques prenant place dans cette séance) et d'une régression logistique pour prédire la réponse thérapeutique. Cette approche a été appliquée à un ensemble de données contenant des transcriptions textuelles de séances de thérapie immersive de 18 patients, et la performance du modèle a été testée sur un ensemble distinct de 17 participants. Le modèle prédictif a correctement prédit les résultats cliniques pour 15 des 17 participants (Hudon et al., 2023).

Il s’agit, à notre connaissance, de la première étude qui combine un algorithme de classification automatisée d’interactions thérapeutique à un algorithme prédictif dans un contexte de psychothérapie. Lors de prédictions d’issues clinique en utilisant un paradigme catégoriel (par exemple, des thèmes bien définis d’interactions thérapeutiques), plusieurs études issues des sciences de la santé rapportent que les modèles mathématiques de type régression logistique (lors d’issues binaires) sont appropriés (Wong et al., 2013). Lorsque l’échantillon de données utilisés pour faire la prédiction est petit, d’autres algorithmes qui utilisent des modalités diverses afin de régulariser les données (par exemple, le modèle de la première étude) sont plus difficilement utilisables car la fiabilité des résultats peut varier selon la dimensionalité des données et la qualité de celles-ci (Emmert-Streib et al, 2019). De plus, pour des échantillons de textes comparables à la base de données TA, la régression logistique peut performer aussi bien que des algorithmes beaucoup plus poussés sur le plan de la complexité des algorithmes (Shyrokykh et al., 2023).

Un aspect important à considérer dans cette étude est la généralisation des données. La combinaison de modèles de classification et de régression a été entrainé avec les données de 18 patients ayant suivi la TA. Les 17 patients avec lesquels les données de leur première séance immersive ont été utilisés pour effectuer la prédiction n’étaient pas connus de la base de données. Malgré, la performance de prédire adéquatement le devenir clinique de 15 de ces 17 patients, il faut comprendre que les 17 patients s’inscrivaient dans le même essai clinique que les 18 patients de la base de données. En ce sens, les critères d’exclusion et inclusion des patients dans cet essai étant très similaire, il est possible que leur profil clinique s’aligne très bien avec les données des patients issus de la base de données, ce qui en expliquerait la performance de l’algorithme. Il est donc essentiel de prendre en compte le contexte clinique très spécifique de la TA et des patients qui la suivent actuellement. Ce modèle ne peut donc pas être prétendu portable à tout type de patients présentant une SRT dans le cadre d’une thérapie similaire. Une validation externe plus poussée aurait pu inclure l’utilisation de données de patients souffrant d’une SRT ayant suivi la TA à l’extérieur de l’IUSMM (Cabitza et al, 2021). Toutefois, les deux seuls autres milieux cliniques offrant la TA n'ont pas de base de données comparatives et leurs thérapies se donnent en anglais dans un contexte culturel très différent du Québec. En ce sens, l'absence d'un groupe comparatif limite les interprétations quant à la validité externe de ce modèle mathématique.

Conclusion de la section

La huitième étude de cette thèse permet d'intégrer pour la première fois un algorithme de classification automatisé à un algorithme de régression logistique dans le contexte de la TA. La performance du modèle est encourageante et met en lumière les principaux défis à explorer afin d'utiliser ce type de prédiction clinique comme aide au processus thérapeutique pour les thérapeutes. Toutefois, dans un cadre spécifique comme la TA et la linéarité des données, les performances du modèle pourrait être un premier pas pour le développement d'une étude plus large afin d'explorer la robustesse de ce modèle et d'en identifier les méthodes d'optimisation. De plus, l'utilisation de variables multimodales pourraient permettre une amélioration des performances du modèles. En ce sens, une exploration plus large des variables à intégrer dans le modèle pourrait être intéressante afin de raffiner la prédiction du devenir clinique de ces patients.

v. Connaître les facteurs prédictifs multimodaux qui pourraient bonifier la prédiction de l'issue clinique des patients

Exploration des émotions prenant place dans la thérapie par Avatar

Les émotions prennent une place importante dans le processus thérapeutique. La régulation émotionnelle, par exemple, aurait une importance par rapport au bénéfice qu'un patient pourrait tirer de sa thérapie (Ehrenreich et al., 2007). C'est dans ce contexte, qu'à la recherche de nouvelles variables à intégrer dans le modèle de la huitième étude que cette neuvième étude a été mise sur pied. L'étude visait à explorer les expressions émotionnelles de patients atteints de SRT suivant une TA, en analysant les transcriptions de sessions immersives et des enregistrements audio de 16 participants. Utilisant une technique d'analyse de contenu avec catégorisation itérative, diverses émotions exprimées par les patients et leurs avatars ont été identifiées et classifiées. Les résultats ont révélé que les patients exprimaient principalement des émotions neutres, de joie et de colère, tandis que les avatars manifestaient surtout de l'intérêt, du dégoût/mépris et des émotions neutres (Hudon et al., 2023).

L'exploration des émotions dans le contexte de la psychothérapie est importante puisqu'elle peut permettre de mieux comprendre le vécu émotionnel du patient lors de celles-ci et faire émerger des points de discussion afin de faire cheminer le patient à travers son vécu. Dans le domaine de la psychiatrie, certaines études rapportent que les émotions peuvent être utile pour prédire le devenir clinique des patients. Par exemple, l'équipe de Lazarus et al. (2021) ont examiné le rôle prédictif de la différenciation des émotions dans les résultats de la psychothérapie pour les patients souffrant de troubles de l'humeur et d'anxiété. Les participants ont complété des évaluations quotidiennes de leurs émotions avant de recevoir une TCC. Les résultats ont révélé qu'une différenciation émotionnelle négative plus élevée, en particulier dans le contexte d'une variabilité émotionnelle plus faible, était associée à une amélioration auto-déclarée plus importante des symptômes de dépression et de stress, mais pas d'anxiété (Lazarus et al., 2021). Cela suggère que les patients capables d'identifier plus distinctement leurs émotions négatives peuvent bénéficier davantage de la psychothérapie, en particulier si leurs expériences émotionnelles sont moins variables. En ce sens, si les thérapeutes étaient davantage en mesure d'aider les patients à identifier ces émotions négatives, cela pourrait les aider à bénéficier des avantages de la thérapie. En schizophrénie, il est reconnu depuis presque trois décades que l'expression des émotions est un des facteurs de prédictions de la rechute chez les patients et qu'il est donc important de mieux comprendre les émotions des patients souffrant de cette maladie (Butzlaff et al., 1998).

La reconnaissance des émotions étant variable chez les patients atteints de schizophrénie, une identification systématique de celles-ci à l'aide de l'apprentissage machine pourrait aider les cliniciens à mieux comprendre les trajectoires cliniques de leurs patients et potentiellement les aider à personnaliser leurs approches afin d'en améliorer les issues cliniques (Zierhut et al., 2022). Une revue de la littérature sur le sujet soulève notamment l'implication de la reconnaissance faciale pour mieux aider les cliniciens à comprendre le vécu émotionnel de leurs patients (Gao et al., 2021). Cette neuvième étude permet donc de mieux comprendre les émotions qui prennent place dans la TA. Il serait intéressant d'explorer leurs variations à travers les séances immersives et d'identifier s'il y a des liens entre ces variations et la réponse des patients à la TA.

Les dyades thérapeutiques issues de la thérapie par Avatar

La dixième et dernière étude de cette thèse a permis de mettre en évidence les dyades d'interactions thérapeutiques les plus prévalentes dans la TA. Un total de 256 verbatims issus de 32 patients ayant suivi une TA a été analysé, identifiant cinq interactions dyadiques se produisant plus fréquemment lors des séances de thérapie. Les résultats suggèrent que ces interactions, telles que "Avatar : Renforcement Patient : Auto-affirmation" et "Avatar : Provocation Patient : Auto-affirmation", jouent un rôle significatif dans le processus thérapeutique de l'AT pour les patients TRS.

L'alliance thérapeutique, qui est la relation de collaboration entre le patient et le thérapeute, est un facteur clé dans l'adhésion au traitement et la réponse à la psychothérapie (Ardito et al., 2011). En ce sens, l'exploration des interactions sur une base dyadique plutôt qu'individuelle permettrait d'approfondir davantage les aspects de la communication de cette dyade dans la TA et leurs relations avec la réponse clinique des patients. Les études de Soma et al. (2020) ont d'ailleurs mis en évidence que de part les interactions prenant place durant la psychothérapie, il y a une dyade émotionnelle qui s'installe entre le patient et le thérapeute. Le patient deviendrait plus labile à travers les séances (expressivité émotionnelle) tandis que le thérapeute serait moins expressif sur le plan émotionnel (Soma et al., 2020). Dans le cadre de la TA ceci demeure à explorer puisque le thérapeute anime l'avatar et ses émotions. À cet effet, le thérapeute peut avoir connaissance des émotions qu'il fait exprimer à l'avatar et donc prendre conscience de leurs effets sur la dyade thérapeutique.

Une autre étude importante portant sur les dyades patients-thérapeutes sur le sujet des dyades dans le domaine de la psychothérapie a mis en évidence trois observations quant aux dyades thérapeutiques : la cohérence du thérapeute est positivement corrélée à une plus grande amélioration des symptômes des clients, en particulier dans les thérapies plus longues, la cohérence du patient et les dynamiques d'influence mutuelle ne se rapportent pas individuellement à l'amélioration des symptôme et l'influence mutuelle (patient envers le thérapeute et thérapeute envers le patient) présente une plus grande amélioration des patients dans les thérapies plus courtes (Li, 2022).

L'influence entre l'avatar et le patient sur le plan des interactions thérapeutiques sera donc intéressant à explorer davantage puisque celle-ci pourrait avoir un lien avec l'amélioration patients pour une thérapie courte comme la TA. La dixième étude permet donc déjà d'identifier une prévalence plus élevée de cinq dyades de façon générale à travers les séances thérapeutiques.

Conclusion de la section

La neuvième et la dixième étude ont présentés des avenues exploratoires intéressantes quant aux émotions et aux dyades d'interactions thérapeutiques prenant place dans la TA. Dans le contexte de thérapies relationnelles comme la TA, il est démontré à travers ces études qu'il y a des différences entre les émotions exprimées par le patient et celles retrouvées dans la personnification de l'avatar. En ce sens, il pourrait être intéressant d'approfondir ces variations dans le temps à travers les séances immersives. De plus, certaines dyades thérapeutiques semblent revenir plus souvent lors de la TA. Il serait intéressant de mieux comprendre l'origine de ces dyades et d'en comprendre leur association (ou non) avec la réponse clinique. Ces deux sujets pourraient être explorés dans des études subséquentes à titre de variables multimodales à intégrer dans le modèle prédictif de la réponse thérapeutique afin de mieux guider les thérapeutes à personnaliser leurs séances de immersives avec leurs patients.

Limites des études

Les études inclues dans cette thèse ont plusieurs limites et celles-ci ont été mentionnés dans chacun des articles. Parmi celles-ci, nous retrouvons le nombre de données, la généralisation des données qui est actuellement limité au contexte de la TA et l'absence de base de données comparatives. À travers les différentes études, il est important de soulignant le caractère évolutif de la base de données de la TA à l'IUSMM. À ce jour, la base de données demeure petite et cela peut exercer une influence sur la performance des modèles mathématiques utilisés et la généralisation des données. Le contexte du Québec (français québécois) limite également la possibilité d'utiliser des dictionnaires standardisés pour optimiser les algorithmes dans la vectorisation des interactions thérapeutiques.

Chapitre 5 – CONCLUSION

Cette thèse a permis d'explorer des pistes d'intégration de l'intelligence artificielle dans la TA pour la prédiction de la réponse clinique des patients souffrant d'un SRT en permettant au thérapeute de bénéficier d'une prédiction basée sur les interactions thérapeutiques prenant place dans la TA. Les modèles prédictifs sont utilisables chez les patients avec des troubles mentaux graves et persistants, mais comportent plusieurs limites. À cet effet, les revues de littératures initiales et les comparaisons entre les algorithmes identifiées pour des bases de données similaires à celle issue de la TA ont permis d'optimiser le LSVC pour la classification automatisée des interactions thérapeutiques prenant place dans la TA. Ces interactions, utilisés comme variables indépendantes dans un modèle de régression logistiques ont permis de prédire la réponse thérapeutique de 15 patients (sur un échantillon de 17) qui n'étaient pas connus de l'algorithme. Ces résultats, bien que préliminaires, permettent de mettre les assises en place pour le développement d'algorithmes plus robustes. Des approches multimodales, incluant les émotions et l'alliance thérapeutique pourraient être utilisés afin de bonifier la réponse clinique et d'avoir davantage une vision quantitative des processus thérapeutiques prenant place dans la TA.

Recherches futures

La réalisation de ces études a donc permis la mise en place d'un projet pilote portant sur l'aide au thérapeute afin de personnaliser la TA en étant guidé par l'intelligence artificielle. Ce projet implique le recrutement de dix patients atteints de SRT et qui ont des hallucinations auditives réfractaires à la médication.

À la fin des séances immersives de la TA, le devenir du patient (une diminution de 20% prédite au PSYRATS-HA ou non) est rendu disponible au thérapeute. En tentant de prédire le résultat de la thérapie, on permettrait d'avoir un aperçu du processus thérapeutique intrinsèque durant les séances de TA et de mettre en évidence ce processus auprès du thérapeute permettra de personnaliser davantage les séances de thérapie. Plus particulièrement, un rapport des séances immersives sera rendu disponible au thérapeute entre les séances immersives qui mettra l’accent sur les interactions thérapeutiques à prioriser et celles à éviter. Deuxièmement, dans cette ère de médecine personnalisée, ce sera une première en termes d'adaptation d'un processus thérapeutique pour guider le patient vers une réponse potentiellement bonne à l’aide de l’intelligence artificielle. Ainsi, un patient ne répondant pas à la TA pourrait éventuellement devenir un bon répondeur avec cette approche. Cela ouvrira la porte à d'autres études sur son applicabilité à d'autres thérapies telles que la TCC.

Références bibliographiques

Albakri, G., Bouaziz, R., Alharthi, W., Kammoun, S., Al-Sarem, M., Saeed, F., & Hadwan, M. (2022). Phobia exposure therapy using virtual and augmented reality: A systematic review. *Applied Sciences*, 12(3), 1672. https://doi.org/10.3390/app12031672

Abdel-Baki, A., & Nicole, L. (2001). Schizophrénie et psychothérapies cognitivo-comportementales [Schizophrenia and cognitive behavioral psychotherapy]. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *46*(6), 511–521. https://doi.org/10.1177/070674370104600605

Adams, R. A., Huys, Q. J., & Roiser, J. P. (2016). Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of neurology, neurosurgery, and psychiatry*, *87*(1), 53–63. https://doi.org/10.1136/jnnp-2015-310737

Alderson-Day, B., Woods, A., Moseley, P., Common, S., Deamer, F., Dodgson, G., & Fernyhough, C. (2020). Voice-hearing and personification: Characterizing social qualities of auditory verbal hallucinations in early psychosis. *Schizophrenia Bulletin*, 47(1), 228–236. https://doi.org/10.1093/schbul/sbaa095

Aleman, A., Kahn, R. S., & Selten, J. P. (2003). Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Archives of general psychiatry*, *60*(6), 565–571. https://doi.org/10.1001/archpsyc.60.6.565

American Psychiatric Association. (2022). Neurodevelopmental disorders. In *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.).

An der Heiden, W., Leber, A., & Häfner, H. (2016). Negative symptoms and their association with depressive symptoms in the long-term course of schizophrenia. *European archives of* *psychiatry and clinical neuroscience*, *266*(5), 387–396. https://doi.org/10.1007/s00406-016-0697-2

Ardito, R. B., & Rabellino, D. (2011). Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. *Frontiers in psychology*, 2, 270. https://doi.org/10.3389/fpsyg.2011.00270

Avasthi, A., Sahoo, S., & Grover, S. (2020). Clinical Practice Guidelines for Cognitive Behavioral Therapy for Psychotic Disorders. *Indian journal of psychiatry*, *62*(Suppl 2), S251–S262. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_774_19

Baker, S. C., Konova, A. B., Daw, N. D., & Horga, G. (2019). A distinct inferential mechanism for delusions in schizophrenia. *Brain : a journal of neurology*, *142*(6), 1797–1812. https://doi.org/10.1093/brain/awz051

Ballesteros, J., Moreno-Calvete, M. C., Santos-Zorrozúa, B., & González-Fraile, E. (2023). Cognitive behavioural therapy plus standard care versus standard care for persistent aggressive behaviour or agitation in people with schizophrenia. *The Cochrane database of systematic reviews*, *7*(7), CD013511. https://doi.org/10.1002/14651858.CD013511.pub2

Bassett, A. S., & Chow, E. W. (2008). Schizophrenia and 22q11.2 deletion syndrome. *Current psychiatry reports*, 10(2), 148–157. https://doi.org/10.1007/s11920-008-0026-1

Beaudoin, M., Potvin, S., Machalani, A., Dellazizzo, L., Bourguignon, L., Phraxayavong, K., & Dumais, A. (2021). The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. *Clinical psychology & psychotherapy*, *28*(3), 500–518. https://doi.org/10.1002/cpp.2556

Beaudoin, M., Hudon, A., Giguère, C. E., Potvin, S., & Dumais, A. (2022). Prediction of quality of life in schizophrenia using machine learning models on data from Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial. *Schizophrenia (Heidelberg, Germany)*, *8*(1), 29. https://doi.org/10.1038/s41537-022-00236-w

Beck, K., McCutcheon, R., Stephenson, L., Schilderman, M., Patel, N., Ramsay, R., & Howes, O. D. (2019). Prevalence of treatment-resistant psychoses in the community: A naturalistic study. *Journal of psychopharmacology (Oxford, England)*, 33(10), 1248–1253. https://doi.org/10.1177/0269881119855995

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(32), 15849–15854. https://doi.org/10.1073/pnas.1903070116

Brisch, R., Saniotis, A., Wolf, R., Bielau, H., Bernstein, H. G., Steiner, J., Bogerts, B., Braun, K., Jankowski, Z., Kumaratilake, J., Henneberg, M., & Gos, T. (2014). The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Frontiers in psychiatry*, *5*, 47. https://doi.org/10.3389/fpsyt.2014.00047

Butzlaff, R. L., & Hooley, J. M. (1998). Expressed emotion and psychiatric relapse: a meta-analysis. *Archives of general psychiatry*, 55(6), 547–552. https://doi.org/10.1001/archpsyc.55.6.547

Cabitza, F., Campagner, A., Soares, F., García de Guadiana-Romualdo, L., Challa, F., Sulejmani, A., Seghezzi, M., & Carobene, A. (2021). The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Computer methods and programs in biomedicine*, 208, 106288. https://doi.org/10.1016/j.cmpb.2021.106288

Canadian Community Health Survey – Mental Health (CCHS – MH). (2012). Percentage of the household population aged 12 and older living in the 10 provinces that met criteria for at least one of six mental disorders (including mood disorders, generalized anxiety disorder, and substance use disorders). Repéré le 19 décembre 2023 à [https://health-infobase.canada.ca/datalab/mental-illness-blog.html]

Cao, Y., Geddes, T. A., Yang, J. Y., & Yang, P. (2020). Ensemble deep learning in bioinformatics. *Nature Machine Intelligence*, 2(9), 500–508. https://doi.org/10.1038/s42256-020-0217-y

Cearns, M., Hahn, T., & Baune, B. T. (2019). Recommendations and future directions for supervised machine learning in psychiatry. *Translational psychiatry*, *9*(1), 271. https://doi.org/10.1038/s41398-019-0607-2

Češková, E., & Šilhán, P. (2021). From Personalized Medicine to Precision Psychiatry?. *Neuropsychiatric disease and treatment*, *17*, 3663–3668. https://doi.org/10.2147/NDT.S337814

Chakrabarti S. (2021). Clozapine resistant schizophrenia: Newer avenues of management. *World journal of psychiatry*, *11*(8), 429–448. https://doi.org/10.5498/wjp.v11.i8.429

Chandler, C., Foltz, P. W., & Elvevåg, B. (2020). Using Machine Learning in Psychiatry: The Need to Establish a Framework That Nurtures Trustworthiness. *Schizophrenia bulletin*, *46*(1), 11–14. https://doi.org/10.1093/schbul/sbz105

Chiu, Y. H., Hsu, C. Y., Lu, M. L., & Chen, C. H. (2020). Augmentation Strategies for Clozapine-Resistant Patients with Schizophrenia. *Current pharmaceutical design*, *26*(2), 218–227. https://doi.org/10.2174/1381612826666200110102254

Chong, H. Y., Teoh, S. L., Wu, D. B., Kotirum, S., Chiou, C. F., & Chaiyakunapruk, N. (2016). Global economic burden of schizophrenia: a systematic review. *Neuropsychiatric disease and treatment*, 12, 357–373. https://doi.org/10.2147/NDT.S96649

Chung, J., & Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain informatics*, 10(1), 1. https://doi.org/10.1186/s40708-022-00180-6

Cieślik, B., Mazurek, J., Rutkowski, S., Kiper, P., Turolla, A., & Szczepańska-Gieracha, J. (2020). Virtual reality in psychiatric disorders: A systematic review of reviews. *Complementary therapies in medicine*, *52*, 102480. https://doi.org/10.1016/j.ctim.2020.102480

Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature. *Front Psychol*, 9, 2086. https://doi.org/10.3389/fpsyg.2018.02086

Cook, S. C., Schwartz, A. C., & Kaslow, N. J. (2017). Evidence-Based Psychotherapy: Advantages and Challenges. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, 14(3), 537–545. https://doi.org/10.1007/s13311-017-0549-4

Correll, C. U., & Schooler, N. R. (2020). Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. *Neuropsychiatric disease and treatment*, *16*, 519–534. https://doi.org/10.2147/NDT.S225643

Correll, C. U., Martin, A., Patel, C., Benson, C., Goulding, R., Kern-Sliwa, J., Joshi, K., Schiller, E., & Kim, E. (2022). Systematic literature review of schizophrenia clinical practice guidelines on acute and maintenance management with antipsychotics. *Schizophrenia (Heidelberg, Germany)*, *8*(1), 5. https://doi.org/10.1038/s41537-021-00192-x

Craig, T. K., Rus-Calafell, M., Ward, T., Leff, J. P., Huckvale, M., Howarth, E., Emsley, R., & Garety, P. A. (2018). AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. *The lancet. Psychiatry*, *5*(1), 31–40. https://doi.org/10.1016/S2215-0366(17)30427-3

Crema, C., Attardi, G., Sartiano, D., & Redolfi, A. (2022). Natural language processing in clinical neuroscience and psychiatry: A review. *Frontiers in psychiatry*, *13*, 946387. https://doi.org/10.3389/fpsyt.2022.946387

Dellazizzo, L., Percie du Sert, O., Phraxayavong, K., Potvin, S., O'Connor, K., & Dumais, A. (2018). Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. *Clinical psychology & psychotherapy*, *25*(6), 878–885. https://doi.org/10.1002/cpp.2322

Dellazizzo, L., Potvin, S., Beaudoin, M., Luigi, M., Dou, B. Y., Giguère, C. É., & Dumais, A. (2019). Cannabis use and violence in patients with severe mental illnesses: A meta-analytical investigation. *Psychiatry research*, 274, 42–48. https://doi.org/10.1016/j.psychres.2019.02.010

Dellazizzo, L., Potvin, S., Luigi, M., & Dumais, A. (2020). Evidence on Virtual Reality-Based Therapies for Psychiatric Disorders: Meta-Review of Meta-Analyses. *Journal of medical Internet research*, *22*(8), e20889. https://doi.org/10.2196/20889

Dellazizzo, L., Potvin, S., Phraxayavong, K., & Dumais, A. (2021). One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. *NPJ schizophrenia*, *7*(1), 9. https://doi.org/10.1038/s41537-021-00139-2

Dellazizzo, L., Giguère, S., Léveillé, N., Potvin, S., & Dumais, A. (2022). A systematic review of relational-based therapies for the treatment of auditory hallucinations in patients with psychotic disorders. *Psychological medicine*, *52*(11), 2001–2008. https://doi.org/10.1017/S003329172200143X

Demšar, J., & Zupan, B. (2021). Hands-on training about overfitting. *PLoS computational biology*, *17*(3), e1008671. https://doi.org/10.1371/journal.pcbi.1008671

Dennison, C. A., Legge, S. E., Pardiñas, A. F., & Walters, J. T. R. (2020). Genome-wide association studies in schizophrenia: Recent advances, challenges and future perspective. *Schizophrenia research*, *217*, 4–12. https://doi.org/10.1016/j.schres.2019.10.048

Deo R. C. (2015). Machine Learning in Medicine. *Circulation*, 132(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Diniz, E., Fonseca, L., Rocha, D., Trevizol, A., Cerqueira, R., Ortiz, B., Brunoni, A. R., Bressan, R., Correll, C. U., & Gadelha, A. (2023). Treatment Resistance in Schizophrenia: a Meta-Analysis of Prevalences and Correlates. *Revista brasileira de psiquiatria (Sao Paulo, Brazil : 1999)*, 10.47626/1516-4446-2023-3126. Advance online publication. https://doi.org/10.47626/1516-4446-2023-3126

Doan, S., Conway, M., Phuong, T. M., & Ohno-Machado, L. (2014). Natural language processing in biomedicine: a unified system architecture overview. *Methods in molecular biology (Clifton, N.J.)*, *1168*, 275–294. https://doi.org/10.1007/978-1-4939-0847-9_16

Dokucu M. E. (2015). Neuromodulation Treatments for Schizophrenia. *Current treatment options in psychiatry*, *2*(3), 339–348. https://doi.org/10.1007/s40501-015-0055-4

du Sert, O. P., Potvin, S., Lipp, O., Dellazizzo, L., Laurelli, M., Breton, R., Lalonde, P., Phraxayavong, K., O'Connor, K., Pelletier, J. F., Boukhalfi, T., Renaud, P., & Dumais, A. (2018). Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. *Schizophrenia research*, *197*, 176–181. https://doi.org/10.1016/j.schres.2018.02.031

Durstewitz, D., Koppe, G., & Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. *Molecular psychiatry*, 24(11), 1583–1598. https://doi.org/10.1038/s41380-019-0365-9

Eckhardt, C. M., Madjarova, S. J., Williams, R. J., Ollivier, M., Karlsson, J., Pareek, A., & Nwachukwu, B. U. (2023). Unsupervised machine learning methods and emerging applications in healthcare. *Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA*, 31(2), 376–381. https://doi.org/10.1007/s00167-022-07233-7

Ehrenreich, J. T., Fairholme, C. P., Buzzella, B. A., Ellard, K. K., & Barlow, D. H. (2007). The Role of Emotion in Psychological Therapy. *Clinical psychology : a publication of the Division of Clinical Psychology of the American Psychological Association*, 14(4), 422–428. https://doi.org/10.1111/j.1468-2850.2007.00102.x

Emmert-Streib, F., & Dehmer, M. (2019). High-dimensional lasso-based computational regression models: Regularization, shrinkage, and selection. *Machine Learning and Knowledge Extraction*, 1(1), 359–383. https://doi.org/10.3390/make1010021

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models With Big Data. *Frontiers in artificial intelligence*, 3, 4. https://doi.org/10.3389/frai.2020.00004

Emsley, R., Chiliza, B., & Schoeman, R. (2008). Predictors of long-term outcome in schizophrenia. *Current opinion in psychiatry*, *21*(2), 173–177. https://doi.org/10.1097/YCO.0b013e3282f33f76

Estave, P. M., Spodnick, M. B., & Karkhanis, A. N. (2022). KOR Control over Addiction Processing: An Exploration of the Mesolimbic Dopamine Pathway. *Handbook of experimental pharmacology*, *271*, 351–377. https://doi.org/10.1007/164_2020_421

Fakhoury M. (2019). Artificial Intelligence in Psychiatry. *Advances in experimental medicine and biology*, *1192*, 119–125. https://doi.org/10.1007/978-981-32-9721-0_6

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *47*(1), 72–89. https://doi.org/10.1038/s41386-021-01132-0

Gao, Z., Zhao, W., Liu, S., Liu, Z., Yang, C., & Xu, Y. (2021). Facial Emotion Recognition in Schizophrenia. *Frontiers in psychiatry*, 12, 633717. https://doi.org/10.3389/fpsyt.2021.633717

George, M., Maheshwari, S., Chandran, S., Manohar, J. S., & Sathyanarayana Rao, T. S. (2017). Understanding the schizophrenia prodrome. *Indian journal of psychiatry*, *59*(4), 505–509. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_464_17

Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., & Ranganath, R. (2020). A Review of Challenges and Opportunities in Machine Learning for Health. AMIA Joint Summits on Translational Science proceedings. *AMIA Joint Summits on Translational Science*, 2020, 191–200.

Gipps, R. G., & Fulford, K. W. (2004). Understanding the clinical concept of delusion: from an estranged to an engaged epistemology. *International review of psychiatry (Abingdon, England)*, *16*(3), 225–235. https://doi.org/10.1080/09540260400003966

Glazer W. M. (2000). Extrapyramidal side effects, tardive dyskinesia, and the concept of atypicality. *The Journal of clinical psychiatry*, *61 Suppl 3*, 16–21.

Goodnick, P. J., Rodriguez, L., & Santana, O. (2002). Antipsychotics: impact on prolactin levels. *Expert opinion on pharmacotherapy*, *3*(10), 1381–1391. https://doi.org/10.1517/14656566.3.10.1381

Grattan D. R. (2015). 60 YEARS OF NEUROENDOCRINOLOGY: The hypothalamo-prolactin axis. *The Journal of endocrinology*, *226*(2), T101–T122. https://doi.org/10.1530/JOE-15-0213

Hamilton, D. F., Ghert, M., & Simpson, A. H. (2015). Interpreting regression models in clinical outcome studies. *Bone & joint research*, 4(9), 152–153. https://doi.org/10.1302/2046-3758.49.2000571

Hayward, M., Overton, J., Dorey, T., & Denney, J. (2009). Relating therapy for people who hear voices: a case series. *Clinical psychology & psychotherapy*, *16*(3), 216–227. https://doi.org/10.1002/cpp.615

Health Quality Ontario (2018). Cognitive Behavioural Therapy for Psychosis: A Health Technology Assessment. *Ontario health technology assessment series*, *18*(5), 1–141.

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., Nordentoft, M., & Glenthøj, B. (2018). Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological psychiatry*, *83*(6), 492–498. https://doi.org/10.1016/j.biopsych.2017.08.017

Hill, C. E., Chui, H., & Baumann, E. (2013). Revisiting and reenvisioning the outcome problem in psychotherapy: an argument to include individualized and qualitative measurement.

Psychotherapy (Chicago, Ill.), 50(1), 68–76. https://doi.org/10.1037/a0030571

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive therapy and research*, *36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Howes, O., McCutcheon, R., & Stone, J. (2015). Glutamate and dopamine in schizophrenia: an update for the 21st century. *Journal of psychopharmacology (Oxford, England)*, *29*(2), 97–115. https://doi.org/10.1177/0269881114563634

Hu, Y., & Gan, H. (2022). Predictive Analysis of Hospital HIS System Usage Satisfaction Based on Machine Learning. *Computational and mathematical methods in medicine*, 2022, 1366407. https://doi.org/10.1155/2022/1366407

Hudon, A., Beaudoin, M., Phraxayavong, K., Dellazizzo, L., Potvin, S., & Dumais, A. (2021). Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms. *JMIR mental health*, 8(10), e22651. https://doi.org/10.2196/22651

Hudon, A., Beaudoin, M., Phraxayavong, K., Dellazizzo, L., Potvin, S., & Dumais, A. (2022). Implementation of a machine learning algorithm for automated thematic annotations in avatar: A linear support vector classifier approach. *Health informatics journal*, 28(4), 14604582221142442. https://doi.org/10.1177/14604582221142442

Hudon, A., Dellazizzo, L., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Association Between Cannabis and Violence in Community-Dwelling Patients With Severe Mental Disorders: A Cross-sectional Study Using Machine Learning. *The Journal of nervous and mental disease*, 211(2), 88–94. https://doi.org/10.1097/NMD.0000000000001604

Hudon, A., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Comparing the performance of machine learning algorithms in the automatic classification of Psychotherapeutic Interactions in avatar therapy. *Machine Learning and Knowledge Extraction*, 5(3), 1119–1130. https://doi.org/10.3390/make5030057

Hudon, A., Beaudoin, M., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Unsupervised Machine Learning Driven Analysis of Verbatims of Treatment-Resistant Schizophrenia Patients Having Followed Avatar Therapy. *Journal of personalized medicine*, 13(5), 801. https://doi.org/10.3390/jpm13050801

Hudon, A., Lammatteo, V., Rodrigues-Coutlée, S., Dellazizzo, L., Giguère, S., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Exploration of the role of emotional expression of treatment-resistant schizophrenia patients having followed virtual reality therapy: a content analysis. *BMC psychiatry*, 23(1), 420. https://doi.org/10.1186/s12888-023-04861-2

Hudon, A., Couture, J., Dellazizzo, L., Beaudoin, M., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Dyadic Interactions of Treatment-Resistant Schizophrenia Patients Having Followed Virtual Reality Therapy: A Content Analysis. *Journal of clinical medicine*, 12(6), 2299. https://doi.org/10.3390/jcm12062299

Hudon, A., Aird, M., & La Haye-Caty, N. (2023). Deciphering the mosaic of therapeutic potential: A scoping review of neural network applications in psychotherapy enhancements. *BioMedInformatics*, 3(4), 1101–1111. https://doi.org/10.3390/biomedinformatics3040066

Hudon, A., Beaudoin, M., Phraxayavong, K., Potvin, S., & Dumais, A. (2023). Enhancing Predictive Power: Integrating a Linear Support Vector Classifier with Logistic Regression for Patient Outcome Prognosis in Virtual Reality Therapy for Treatment-Resistant Schizophrenia.

Journal of personalized medicine, 13(12), 1660. https://doi.org/10.3390/jpm13121660

Huang, Y. C., Lee, Y., Lee, C. Y., Lin, P. Y., Hung, C. F., Lee, S. Y., & Wang, L. J. (2020). Defining cognitive and functional profiles in schizophrenia and affective disorders. *BMC psychiatry*, 20(1), 39. https://doi.org/10.1186/s12888-020-2459-y

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior therapy*, *51*(5), 675–687. https://doi.org/10.1016/j.beth.2020.05.002

Jin, K. W., Li, Q., Xie, Y., & Xiao, G. (2023). Artificial intelligence in mental healthcare: an overview and future perspectives. *The British journal of radiology*, *96*(1150), 20230213. https://doi.org/10.1259/bjr.20230213

Joodaki, H., Gepner, B., & Kerrigan, J. (2021). Leveraging machine learning for predicting human body model response in restraint design simulations. *Computer methods in biomechanics and biomedical engineering*, *24*(6), 597–611. https://doi.org/10.1080/10255842.2020.1841754

Kart, A., Özdel, K., & Türkçapar, M. H. (2021). Cognitive Behavioral Therapy in Treatment of Schizophrenia. *Noro psikiyatri arsivi*, *58*(Suppl 1), S61–S65. https://doi.org/10.29399/npa.27418

Kay, S. R., & Murrill, L. M. (1990). Predicting outcome of schizophrenia: significance of symptom profiles and outcome dimensions. *Comprehensive psychiatry*, *31*(2), 91–102. https://doi.org/10.1016/0010-440x(90)90012-h

Keefe, R. S., Eesley, C. E., & Poe, M. P. (2005). Defining a cognitive function decrement in schizophrenia. *Biological psychiatry*, *57*(6), 688–691. https://doi.org/10.1016/j.biopsych.2005.01.003

Kern, A. C., & Ellermeier, W. (2020). Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence. *Front Robot AI*, 7, 20. https://doi.org/10.3389/frobt.2020.00020

Khanbhai, M., Warren, L., Symons, J., Flott, K., Harrison-White, S., Manton, D., Darzi, A., & Mayer, E. (2022). Using natural language processing to understand, facilitate and maintain continuity in patient experience across transitions of care. *International journal of medical informatics*, *157*, 104642. https://doi.org/10.1016/j.ijmedinf.2021.104642

Kim, S., & Kim, E. (2020). The Use of Virtual Reality in Psychiatry: A Review. *Soa--ch'ongsonyon chongsin uihak = Journal of child & adolescent psychiatry*, 31(1), 26–32. https://doi.org/10.5765/jkacap.190037

Kleinberg, J., Ludwig, J., Mullainathan, S., & Raghavan, M. (2023). The Inversion Problem: Why Algorithms Should Infer Mental State and Not Just Predict Behavior. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 17456916231212138. Advance online publication. https://doi.org/10.1177/17456916231212138

Kleine, A. K., Kokje, E., Lermer, E., & Gaube, S. (2023). Attitudes Toward the Adoption of 2 Artificial Intelligence-Enabled Mental Health Tools Among Prospective Psychotherapists: Cross-sectional Study. *JMIR human factors*, *10*, e46859. https://doi.org/10.2196/46859

Kishi, T., Ikuta, T., Matsuda, Y., Sakuma, K., & Iwata, N. (2020). Aripiprazole vs. brexpiprazole for acute schizophrenia: a systematic review and network meta-analysis. *Psychopharmacology*, *237*(5), 1459–1470. https://doi.org/10.1007/s00213-020-05472-5

Kiran, C., & Chaudhury, S. (2009). Understanding delusions. *Industrial psychiatry journal*, *18*(1), 3–18. https://doi.org/10.4103/0972-6748.57851

Kirkpatrick, B., Miller, B., García-Rizo, C., & Fernandez-Egea, E. (2014). Schizophrenia: a systemic disorder. *Clinical schizophrenia & related psychoses*, *8*(2), 73–79. https://doi.org/10.3371/CSRP.KIMI.031513

Kolecki, R., Pręgowska, A., Dąbrowa, J., Skuciński, J., Pulanecki, T., Walecki, P., van Dam, P. M., Dudek, D., Richter, P., & Proniewska, K. (2022). Assessment of the utility of Mixed Reality in medical education. *Translational Research in Anatomy*, 28. https://doi.org/10.1016/j.tria.2022.100214

Koutsouleris, N., Hauser, T. U., Skvortsova, V., & De Choudhury, M. (2022). From promise to practice: towards the realisation of AI-informed mental health care. *The Lancet. Digital health*, *4*(11), e829–e840. https://doi.org/10.1016/S2589-7500(22)00153-4

Krakovska, O., Christie, G., Sixsmith, A., Ester, M., & Moreno, S. (2019). Performance comparison of linear and non-linear feature selection methods for the analysis of large survey datasets. *PloS one*, 14(3), e0213584. https://doi.org/10.1371/journal.pone.0213584

Lally, J., & MacCabe, J. H. (2015). Antipsychotic medication in schizophrenia: a review. *British medical bulletin*, *114*(1), 169–179. https://doi.org/10.1093/bmb/ldv017

Laws, K. R., Darlington, N., Kondel, T. K., McKenna, P. J., & Jauhar, S. (2018). Cognitive Behavioural Therapy for schizophrenia - outcomes for functioning, distress and quality of life: a meta-analysis. *BMC psychology*, *6*(1), 32. https://doi.org/10.1186/s40359-018-0243-2

Lazarus, G., & Fisher, A. J. (2021). Negative Emotion Differentiation Predicts Psychotherapy Outcome: Preliminary Findings. *Frontiers in psychology*, 12, 689407.

https://doi.org/10.3389/fpsyg.2021.689407

Leff, J., Williams, G., Huckvale, M., Arbuthnot, M., & Leff, A. P. (2014). Avatar therapy for persecutory auditory hallucinations: What is it and how does it work?. *Psychosis*, *6*(2), 166–176. https://doi.org/10.1080/17522439.2013.773457

Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *Journal of medical Internet research*, *23*(5), e15708. https://doi.org/10.2196/15708

Leucht, S., Cipriani, A., Spineli, L., Mavridis, D., Orey, D., Richter, F., Samara, M., Barbui, C., Engel, R. R., Geddes, J. R., Kissling, W., Stapf, M. P., Lässig, B., Salanti, G., & Davis, J. M. (2013). Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet (London, England)*, *382*(9896), 951–962. https://doi.org/10.1016/S0140-6736(13)60733-3

Leung L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care*, 4(3), 324–327. https://doi.org/10.4103/2249-4863.161306

Li X. (2022). The "dyadic dance": Exploring therapist-client dynamics and client symptom change using actor-partner interdependence modeling and multilevel mixture modeling. *Journal of counseling psychology*, 69(4), 474–489. https://doi.org/10.1037/cou0000599

Li, R., Ma, X., Wang, G., Yang, J., & Wang, C. (2016). Why sex differences in schizophrenia?. *Journal of translational neuroscience*, *1*(1), 37–42.

Li, X., Zhou, W., & Yi, Z. (2022). A glimpse of gender differences in schizophrenia. *General psychiatry*, *35*(4), e100823. https://doi.org/10.1136/gpsych-2022-100823

Lieberman J. A. (2006). Neurobiology and the natural history of schizophrenia. *The Journal of clinical psychiatry*, *67*(10), e14.

Lin, E., Lin, C. H., & Lane, H. Y. (2021). Applying a bagging ensemble machine learning approach to predict functional outcome of schizophrenia with clinical symptoms and cognitive functions. *Scientific reports*, 11(1), 6922. https://doi.org/10.1038/s41598-021-86382-0

López-Silva, P., de Prado-Gordillo, M. N., & Fernández-Castro, V. (2024). What are delusions? Examining the typology problem. *Wiley interdisciplinary reviews. Cognitive science*, e1674. Advance online publication. https://doi.org/10.1002/wcs.1674

MacCabe, J. H., Brébion, G., Reichenberg, A., Ganguly, T., McKenna, P. J., Murray, R. M., & David, A. S. (2012). Superior intellectual ability in schizophrenia: neuropsychological characteristics. *Neuropsychology*, *26*(2), 181–190. https://doi.org/10.1037/a0026376

Maia, T. V., Huys, Q. J. M., & Frank, M. J. (2017). Theory-Based Computational Psychiatry. *Biological psychiatry*, *82*(6), 382–384. https://doi.org/10.1016/j.biopsych.2017.07.016

Manchia, M., Pisanu, C., Squassina, A., & Carpiniello, B. (2020). Challenges and Future Prospects of Precision Medicine in Psychiatry. *Pharmacogenomics and personalized medicine*, *13*, 127–140. https://doi.org/10.2147/PGPM.S198225

McCutcheon, R. A., Abi-Dargham, A., & Howes, O. D. (2019). Schizophrenia, Dopamine and the Striatum: From Biology to Symptoms. *Trends in neurosciences*, *42*(3), 205–220. https://doi.org/10.1016/j.tins.2018.12.004

McCutcheon, R. A., Krystal, J. H., & Howes, O. D. (2020). Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *19*(1), 15–33. https://doi.org/10.1002/wps.20693

McLachlan, N. M., Phillips, D. S., Rossell, S. L., & Wilson, S. J. (2013). Auditory processing and hallucinations in schizophrenia. *Schizophrenia research*, *150*(2-3), 380–385. https://doi.org/10.1016/j.schres.2013.08.039

Meehan, A. J., Lewis, S. J., Fazel, S., Fusar-Poli, P., Steyerberg, E. W., Stahl, D., & Danese, A. (2022). Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. *Molecular psychiatry*, 27(6), 2700–2708. https://doi.org/10.1038/s41380-022-01528-4

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12), 1321-1329.

Molstrom, I. M., Nordgaard, J., Urfer-Parnas, A., Handest, R., Berge, J., & Henriksen, M. G. (2022). The prognosis of schizophrenia: A systematic review and meta-analysis with meta-regression of 20-year follow-up studies. *Schizophrenia research*, *250*, 152–163. https://doi.org/10.1016/j.schres.2022.11.010

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in cognitive sciences*, *16*(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

Montazeri, M., Montazeri, M., Bahaadinbeigy, K., Montazeri, M., & Afraz, A. (2022). Application of machine learning methods in predicting schizophrenia and bipolar disorders: A systematic review. *Health science reports*, 6(1), e962. https://doi.org/10.1002/hsr2.962

Morrison A. K. (2009). Cognitive behavior therapy for people with schizophrenia. *Psychiatry (Edgmont (Pa. : Township))*, *6*(12), 32–39.

Moscoso, C., Nazari, M., & Matusiak, B. S. (2022). Stereoscopic Images and Virtual Reality techniques in daylighting research: A method-comparison study. *Building and Environment*, 214. https://doi.org/10.1016/j.buildenv.2022.108962

Mosolov, S. N., & Yaltonskaya, P. A. (2022). Primary and Secondary Negative Symptoms in Schizophrenia. *Frontiers in psychiatry*, *12*, 766692. https://doi.org/10.3389/fpsyt.2021.766692

Naderalvojoud, B., & Hernandez-Boussard, T. (2024). Improving machine learning with ensemble learning on observational healthcare data. AMIA ... Annual Symposium proceedings. *AMIA Symposium*, 2023, 521–529.

Nadif, M., & Role, F. (2021). Unsupervised and self-supervised deep learning approaches for biomedical text mining. *Briefings in bioinformatics*, 22(2), 1592–1603. https://doi.org/10.1093/bib/bbab016

Nakao, M., Shirotsuki, K., & Sugaya, N. (2021). Cognitive-behavioral therapy for management of mental health and stress-related disorders: Recent advances in techniques and technologies. *BioPsychoSocial medicine*, *15*(1), 16. https://doi.org/10.1186/s13030-021-00219-w

Naithani, N., Atal, A. T., Tilak, T. V. S. V. G. K., Vasudevan, B., Misra, P., & Sinha, S. (2021). Precision medicine: Uses and challenges. *Medical journal, Armed Forces India*, *77*(3), 258–265. https://doi.org/10.1016/j.mjafi.2021.06.020

Nichols, J. A., Herbert Chan, H. W., & Baker, M. A. B. (2019). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical reviews*, *11*(1), 111–118. https://doi.org/10.1007/s12551-018-0449-9

Nicole, L., Lesage, A., & Lalonde, P. (1992). Lower incidence and increased male:female ratio in schizophrenia. *The British journal of psychiatry : the journal of mental science*, *161*, 556–557. https://doi.org/10.1192/bjp.161.4.556

Nucifora, F. C., Jr, Woznica, E., Lee, B. J., Cascella, N., & Sawa, A. (2019). Treatment resistant schizophrenia: Clinical, biological, and therapeutic perspectives. *Neurobiology of disease*, *131*, 104257. https://doi.org/10.1016/j.nbd.2018.08.016

Ochoa, S., Usall, J., Cobo, J., Labad, X., & Kulkarni, J. (2012). Gender differences in schizophrenia and first-episode psychosis: a comprehensive literature review. *Schizophrenia research and treatment*, *2012*, 916198. https://doi.org/10.1155/2012/916198

Ozomaro, U., Wahlestedt, C., & Nemeroff, C. B. (2013). Personalized medicine in psychiatry: problems and promises. *BMC medicine*, *11*, 132. https://doi.org/10.1186/1741-7015-11-132

Palinkas L. A. (2014). Qualitative and mixed methods in mental health services and implementation research. *Journal of clinical child and adolescent psychology : the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association*, Division 53, 43(6), 851–861. https://doi.org/10.1080/15374416.2014.910791

Park, M. J., Kim, D. J., Lee, U., Na, E. J., & Jeon, H. J. (2019). A Literature Overview of Virtual Reality (VR) in Treatment of Psychiatric Disorders: Recent Advances and Limitations. *Frontiers in psychiatry*, *10*, 505. https://doi.org/10.3389/fpsyt.2019.00505

Patel, K. R., Cherian, J., Gohil, K., & Atkinson, D. (2014). Schizophrenia: overview and treatment options. *P & T : a peer-reviewed journal for formulary management*, *39*(9), 638–645.

Pelin, H., Ising, M., Stein, F., Meinert, S., Meller, T., Brosch, K., Winter, N. R., Krug, A., Leenings, R., Lemke, H., Nenadić, I., Heilmann-Heimbach, S., Forstner, A. J., Nöthen, M. M., Opel, N., Repple, J., Pfarr, J., Ringwald, K., Schmitt, S., Thiel, K., … Andlauer, T. F. M. (2021). Identification of transdiagnostic psychiatric disorder subtypes using unsupervised learning. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *46*(11), 1895–1905. https://doi.org/10.1038/s41386-021-01051-0

Peña, J., Segarra, R., Ojeda, N., García, J., Eguiluz, J. I., & Gutiérrez, M. (2012). Do the same factors predict outcome in schizophrenia and non-schizophrenia syndromes after first-episode psychosis? A two-year follow-up study. *Journal of psychiatric research*, *46*(6), 774–781. https://doi.org/10.1016/j.jpsychires.2012.03.014

Percha, B. (2021). Modern clinical text mining: A guide and review. *Annual Review of Biomedical Data Science*, 4(1), 165–187. https://doi.org/10.1146/annurev-biodatasci-030421-030931

Pham, K. T., Nabizadeh, A., & Selek, S. (2022). Artificial Intelligence and Chatbots in Psychiatry. *The Psychiatric quarterly*, *93*(1), 249–253. https://doi.org/10.1007/s11126-022-09973-8

Pickard B. (2011). Progress in defining the biological causes of schizophrenia. *Expert reviews in molecular medicine*, *13*, e25. https://doi.org/10.1017/S1462399411001955

Potuzak, M., Ravichandran, C., Lewandowski, K. E., Ongür, D., & Cohen, B. M. (2012). Categorical vs dimensional classifications of psychotic disorders. *Comprehensive psychiatry*, *53*(8), 1118–1129. https://doi.org/10.1016/j.comppsych.2012.04.010

Price, G. D., Heinz, M. V., Zhao, D., Nemesure, M., Ruan, F., & Jacobson, N. C. (2022). An unsupervised machine learning approach using passive movement data to understand depression and schizophrenia. *Journal of affective disorders*, *316*, 132–139. https://doi.org/10.1016/j.jad.2022.08.013

Quinlan, E. B., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Bromberg, U., Büchel, C., Desrivières, S., Flor, H., Frouin, V., Garavan, H., Heinz, A., Brühl, R., Martinot, J. L., Paillère Martinot, M. L., Nees, F., Orfanos, D. P., Paus, T., Poustka, L., Hohmann, S., Smolka, M. N., … IMAGEN Consortium (2020). Identifying biological markers for improved precision medicine in psychiatry. *Molecular psychiatry*, *25*(2), 243–253. https://doi.org/10.1038/s41380-019-0555-5

Rathod, S., Phiri, P., & Kingdon, D. (2010). Cognitive behavioral therapy for schizophrenia. *The Psychiatric clinics of North America*, *33*(3), 527–536. https://doi.org/10.1016/j.psc.2010.04.009

Remington, G., Addington, D., Honer, W., Ismail, Z., Raedler, T., & Teehan, M. (2017). Guidelines for the Pharmacotherapy of Schizophrenia in Adults. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *62*(9), 604–616. https://doi.org/10.1177/0706743717720448

Richter, A., Petrovic, A., Diekhof, E. K., Trost, S., Wolter, S., & Gruber, O. (2015). Hyperresponsivity and impaired prefrontal control of the mesolimbic reward system in schizophrenia. *Journal of psychiatric research*, *71*, 8–15. https://doi.org/10.1016/j.jpsychires.2015.09.005

Richter-Levin G. (2004). The amygdala, the hippocampus, and emotional modulation of memory. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, *10*(1), 31–39. https://doi.org/10.1177/1073858403259955

Riley, R. D., Ensor, J., Snell, K. I., Debray, T. P., Altman, D. G., Moons, K. G., & Collins, G. S. (2016). External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ (Clinical research ed.)*, 353, i3140.

https://doi.org/10.1136/bmj.i3140

Robison, A. J., Thakkar, K. N., & Diwadkar, V. A. (2020). Cognition and Reward Circuits in Schizophrenia: Synergistic, Not Separate. *Biological psychiatry*, *87*(3), 204–214. https://doi.org/10.1016/j.biopsych.2019.09.021

Roche, D., & Russell, V. (2021). Can precision medicine advance psychiatry?. *Irish journal of psychological medicine*, *38*(3), 163–168. https://doi.org/10.1017/ipm.2020.79

Rootes-Murdy, K., Goldsmith, D. R., & Turner, J. A. (2022). Clinical and Structural Differences in Delusions Across Diagnoses: A Systematic Review. *Frontiers in integrative neuroscience*, *15*, 726321. https://doi.org/10.3389/fnint.2021.726321

Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *PloS one*, 16(8), e0254937. https://doi.org/10.1371/journal.pone.0254937

Sarker I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN computer science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Scaccia, J. P., & Scott, V. C. (2021). 5335 days of Implementation Science: using natural language processing to examine publication trends and topics. *Implementation science : IS*, *16*(1), 47. https://doi.org/10.1186/s13012-021-01120-4

Schennach-Wolff, R., Jäger, M., Seemüller, F., Obermeier, M., Messer, T., Laux, G., Pfeiffer, H., Naber, D., Schmidt, L. G., Gaebel, W., Huff, W., Heuser, I., Maier, W., Lemke, M. R., Rüther, E., Buchkremer, G., Gastpar, M., Möller, H. J., & Riedel, M. (2009). Defining and predicting functional outcome in schizophrenia and schizophrenia spectrum disorders. *Schizophrenia research*, *113*(2-3), 210–217. https://doi.org/10.1016/j.schres.2009.05.032

Seeman P. (2002). Atypical antipsychotics: mechanism of action. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *47*(1), 27–38.

Shafiei, S. B., Lone, Z., Elsayed, A. S., Hussein, A. A., & Guru, K. A. (2020). Identifying mental health status using deep neural network trained by visual metrics. *Translational psychiatry*, 10(1), 430. https://doi.org/10.1038/s41398-020-01117-5

Shattock, L., Berry, K., Degnan, A., & Edge, D. (2018). Therapeutic alliance in psychological therapy for people with schizophrenia and related psychoses: A systematic review. *Clinical psychology & psychotherapy*, *25*(1), e60–e85. https://doi.org/10.1002/cpp.2135

Shyrokykh, K., Girnyk, M., & Dellmuth, L. (2023). Short text classification with machine learning in the social sciences: The case of climate change on Twitter. *PloS one*, 18(9), e0290762. https://doi.org/10.1371/journal.pone.0290762

Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2. https://doi.org/10.3389/frvir.2021.647997

Soma, C. S., Baucom, B. R. W., Xiao, B., Butner, J. E., Hilpert, P., Narayanan, S., Atkins, D. C., & Imel, Z. E. (2020). Coregulation of therapist and client emotion during psychotherapy. *Psychotherapy research : journal of the Society for Psychotherapy Research*, 30(5), 591–603. https://doi.org/10.1080/10503307.2019.1661541

Stanghellini, G., & Raballo, A. (2015). Differential typology of delusions in major depression and schizophrenia. A critique to the unitary concept of 'psychosis'. *Journal of affective disorders*, *171*, 171–178. https://doi.org/10.1016/j.jad.2014.09.027

Starzer, M. S. K., Nordentoft, M., & Hjorthøj, C. (2018). Rates and Predictors of Conversion to Schizophrenia or Bipolar Disorder Following Substance-Induced Psychosis. *The American* *journal of psychiatry*, 175(4), 343–350. https://doi.org/10.1176/appi.ajp.2017.17020223

Stilo, S. A., & Murray, R. M. (2019). Non-Genetic Factors in Schizophrenia. *Current psychiatry reports*, *21*(10), 100. https://doi.org/10.1007/s11920-019-1091-3

Stip, E., & Tourjman, V. (2010). Aripiprazole in schizophrenia and schizoaffective disorder: A review. *Clinical therapeutics*, *32 Suppl 1*, S3–S20. https://doi.org/10.1016/j.clinthera.2010.01.021

Stroup, T. S., & Gray, N. (2018). Management of common adverse effects of antipsychotic medications. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *17*(3), 341–356. https://doi.org/10.1002/wps.20567

Szymańska, A., Dobrenko, K., & Grzesiuk, L. (2017). Characteristics and experience of the patient in psychotherapy and the psychotherapy's effectiveness. A structural approach. Cechy i doświadczenia pacjenta z przebiegu psychoterapii oraz skuteczność psychoterapii. Podejście strukturalne. *Psychiatria polska*, 51(4), 619–631. https://doi.org/10.12740/PP/62483

Tandon R. (2011). Antipsychotics in the treatment of schizophrenia: an overview. *The Journal of clinical psychiatry*, *72 Suppl 1*, 4–8. https://doi.org/10.4088/JCP.10075su1.01

Tandon, R., Nasrallah, H., Akbarian, S., Carpenter, W. T., Jr, DeLisi, L. E., Gaebel, W., Green, M. F., Gur, R. E., Heckers, S., Kane, J. M., Malaspina, D., Meyer-Lindenberg, A., Murray, R., Owen, M., Smoller, J. W., Yassine, W., & Keshavan, M. (2023). The schizophrenia syndrome, circa 2024: What we know and how that informs its nature. *Schizophrenia research*, *264*, 1–28. Advance online publication. https://doi.org/10.1016/j.schres.2023.11.015

Thara, R., & Kamath, S. (2015). Women and schizophrenia. *Indian journal of psychiatry*, *57*(Suppl

2), S246–S251. https://doi.org/10.4103/0019-5545.161487

Thomas, N., Hayward, M., Peters, E., van der Gaag, M., Bentall, R. P., Jenner, J., Strauss, C., Sommer, I. E., Johns, L. C., Varese, F., García-Montes, J. M., Waters, F., Dodgson, G., & McCarthy-Jones, S. (2014). Psychological therapies for auditory hallucinations (voices): current status and key directions for future research. *Schizophrenia bulletin*, *40 Suppl 4*(Suppl 4), S202–S212. https://doi.org/10.1093/schbul/sbu037

Tonelli, M. R., & Shirts, B. H. (2017). Knowledge for Precision Medicine: Mechanistic Reasoning and Methodological Pluralism. *JAMA*, *318*(17), 1649–1650. https://doi.org/10.1001/jama.2017.11914

Toren, P., Ratner, S., Laor, N., & Weizman, A. (2004). Benefit-risk assessment of atypical antipsychotics in the treatment of schizophrenia and comorbid disorders in children and adolescents. *Drug safety*, *27*(14), 1135–1156. https://doi.org/10.2165/00002018-200427140-00005

Tripathi, A., Kar, S. K., & Shukla, R. (2018). Cognitive Deficits in Schizophrenia: Understanding the Biological Correlates and Remediation Strategies. *Clinical psychopharmacology and neuroscience : the official scientific journal of the Korean College of Neuropsychopharmacology*, *16*(1), 7–17. https://doi.org/10.9758/cpn.2018.16.1.7

Tufail, S. *et al.* (2023) 'Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms', *Electronics*, 12(8), p. 1789. doi:10.3390/electronics12081789.

Turner, D. T., van der Gaag, M., Karyotaki, E., & Cuijpers, P. (2014). Psychological interventions for psychosis: a meta-analysis of comparative outcome studies. *The American journal of psychiatry*, *171*(5), 523–538. https://doi.org/10.1176/appi.ajp.2013.13081159

Turso-Finnich, T., Jensen, R. O., Jensen, L. X., Konge, L., & Thinggaard, E. (2023). Virtual Reality Head-Mounted Displays in Medical Education: A Systematic Review. *Simul Healthc*, 18(1), 42-50. https://doi.org/10.1097/SIH.0000000000000636

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access*, 7, 65579–65615. https://doi.org/10.1109/access.2019.2916648

van Dee, V., Schnack, H. G., & Cahn, W. (2023). Systematic review and meta-analysis on predictors of prognosis in patients with schizophrenia spectrum disorders: An overview of current evidence and a call for prospective research and open access to datasets. *Schizophrenia research*, *254*, 133–142. https://doi.org/10.1016/j.schres.2023.02.024

Verma, S., Goel, T., Tanveer, M., Ding, W., Sharma, R., & Murugan, R. (2023a). Machine learning techniques for the schizophrenia diagnosis: A comprehensive review and future research directions. *Journal of Ambient Intelligence and Humanized Computing*, 14(5), 4795–4807. https://doi.org/10.1007/s12652-023-04536-6

Vergara, D., Rubio, M., & Lorenzo, M. (2017). On the Design of Virtual Reality Learning Environments in Engineering. *Multimodal Technologies and Interaction*, 1(2). https://doi.org/10.3390/mti1020011

Wang, X. J., & Krystal, J. H. (2014). Computational psychiatry. *Neuron*, *84*(3), 638–654. https://doi.org/10.1016/j.neuron.2014.10.018

Wenzel A. (2017). Basic Strategies of Cognitive Behavioral Therapy. *The Psychiatric clinics of North America*, *40*(4), 597–609. https://doi.org/10.1016/j.psc.2017.07.001

Wright, P., & O'Flaherty, L. (2003). Antipsychotic drugs: atypical advantages and typical disadvantages. *Irish journal of psychological medicine*, *20*(1), 24–27. https://doi.org/10.1017/S0790966700007497

Wong, Z. S., & Akiyama, M. (2013). Statistical text classifier to detect specific type of medical incidents. *Studies in health technology and informatics*, 192, 1053.

Xie, J., Wang, Y., Ye, C., Li, X. J., & Lin, L. (2024). Distinctive Patterns of 5-Methylcytosine and 5-Hydroxymethylcytosine in Schizophrenia. *International journal of molecular sciences*, *25*(1), 636. https://doi.org/10.3390/ijms25010636

Yang, Z., Chen, C., Li, H., Yao, L., & Zhao, X. (2020). Unsupervised Classifications of Depression Levels Based on Machine Learning Algorithms Perform Well as Compared to Traditional Norm-Based Classifications. *Frontiers in psychiatry*, *11*, 45. https://doi.org/10.3389/fpsyt.2020.00045

Young, A. T., Amara, D., Bhattacharya, A., & Wei, M. L. (2021). Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *The Lancet. Digital health*, *3*(9), e599–e611. https://doi.org/10.1016/S2589-7500(21)00132-1

Zhang, W., Yoshida, T., & Tang, X. (2008). Text classification based on multi-word with support vector machine. *Knowledge-Based Systems*, 21(8), 879–886. https://doi.org/10.1016/j.knosys.2008.03.044

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. *NPJ digital medicine*, 5(1), 46. https://doi.org/10.1038/s41746-022-00589-7

Zhao, X., Rangaprakash, D., Denney, T. S., Jr, Katz, J. S., Dretsch, M. N., & Deshpande, G. (2018).

Identifying neuropsychiatric disorders using unsupervised clustering methods: Data and code. *Data in brief*, *22*, 570–573. https://doi.org/10.1016/j.dib.2018.01.080

Zierhut, M., Böge, K., Bergmann, N., Hahne, I., Braun, A., Kraft, J., Ta, T. M. T., Ripke, S., Bajbouj, M., & Hahn, E. (2022). The Relationship Between the Recognition of Basic Emotions and Negative Symptoms in Individuals With Schizophrenia Spectrum Disorders - An Exploratory Study. *Frontiers in psychiatry*, 13, 865226. https://doi.org/10.3389/fpsyt.2022.865226

Zipursky R. B. (2014). Why are the outcomes in patients with schizophrenia so poor?. *The Journal of clinical psychiatry*, *75 Suppl 2*, 20–24. https://doi.org/10.4088/JCP.13065su1.05